Homology modelling of rat kallikrein rK9, a member of the tissue kallikrein family: implications for substrate specificity and inhibitor binding

Thierry Moreau¹ and Francis Gauthier

Laboratory of Enzymology and Protein Chemistry, CNRS URA 1334,
University François Rabelais, 2bis Boulevard Tonnellé, F-37032 Tours Cedex, France

¹To whom correspondence should be addressed

The rat kallikrein rK9 is one of the six members of the rat tissue kallikrein family isolated to date. It is 84% identical to rK2 (tonin), and both proteinases are thought to have vasoconstrictive properties. Recently we have shown that rK9 and rK2 have distinct substrate specificities and sensitivities to inhibitors, despite their similar sequences. Unlike all other mammalian kallikrein-related proteinases, rK9 is resistant to inhibition by aprotinin. We have developed a 3-D model of rK9, based on the known X-ray structures of rK2, porcine kallikrein and bovine trypsin, to identify the structural features underlying this functional diversity. The final rK9 model is structurally similar to rK2, but variable regions surrounding the active site differ quite markedly from the reference proteins. The kallikrein loop, which differs from that in porcine kallikrein by a seven-residue insertion, has been generated de novo and subjected to simulated annealing to assess its influence on the restricted substrate specificity of these proteinases. The proposed conformation of the specificity pocket in rK9 differs from that of other serine proteinases, but it can still accommodate both aromatic and basic amino acid side chains at the substrate P₁ position, thus explaining the dual chymotrypsin and trypsin-like activity of rK9. The electrostatic potentials of rK9 and aprotinin were calculated using the finite difference Poisson–Boltzmann method. They indicated a large positive region near the active site of rK9 not found in related proteinases because of positively charged residues at positions 61 and 65 in rK9. They generate a positive region, which overlaps a positive region in aprotinin, and may prevent aprotinin binding. A single mutation in aprotinin is suggested that might allow kallikrein rK9 inhibition by aprotinin. This model contributes significantly to our understanding of the structure–function relationships among proteinases of the tissue kallikrein family.

Keywords: electrostatic potentials/homology modelling/rat kallikrein/serine proteinase inhibitor

Introduction

Six protein products of the rat tissue kallikrein gene family have been identified so far and correlated unambiguously with their corresponding genes (Gauthier et al., 1992). The kinin-releasing enzyme rK1 (the abbreviations for tissue kallikreins follow the nomenclature proposed at the Kinin '91 Munich Symposium; Berg et al., 1992) and rK2 (tonin) were the first to be characterized. The other four, rK7, rK8, rK9 and rK10, have been described more recently but no definite biological function has been attributed to them. rK2 and rK9 may release angiotensin II from angiotensinogen and ANP from pro-ANP-mimicking peptides (Boucher et al., 1974; Moreau et al., 1992, 1995), and it has been suggested that rK10 is a T-kininogenase (Xiong et al., 1990). All the rat kallikreins are synthesized by the submandibular gland, but rK9 is also found in large amounts in the prostate (Windericks et al., 1989). Two of the three kallikreins described to date in humans, hK2 and hK3 (prostatic-specific antigen; PSA), are also mainly present in the prostate (Dubé, 1994). Like rK2 (tonin), which is similar (84% amino acid identity), rK9 has been reported to be a tonin-like enzyme with vasoconstrictive activity (Yamaguchi et al., 1991). Recently we have shown that these two proteinases have similar but distinct substrate specificities and sensitivities to inhibitors, despite their similar sequences (Moreau et al., 1992).

A 3-D structural model of rK9 was developed using homology modelling techniques and the known crystal structures of rK2, porcine kallikrein and bovine trypsin as reference proteins. This model has been used to identify the structural features leading to the above differences. Homology modelling is a valuable tool for obtaining a 3-D structure from a protein sequence, provided that the 3-D structure of at least one homologous protein is known (Ring and Cohen, 1993; Siezen et al., 1994). Comparison of the homologous structures of aspartyl proteinases and of serine proteinases (Tang et al., 1978; Read and James, 1984; Greer, 1990) has led to the current concepts of modelling procedures. This report describes the homology modelling of kallikrein rK9, a serine proteinase of the chymotrypsin family. Comparison of the model with other kallikreins indicates that the highly variable regions correspond to surface loops surrounding the active site. This strongly suggests that structural and sequence differences within these loops account for the diversity of substrate specificities among rat kallikreins. The conformations of the highly flexible ‘kallikrein loop’ of rK9 have been explored using molecular dynamics simulations to assess the influence of this loop on the restricted substrate specificity and sensitivity to inhibitors of kallikreins rK2 and rK9. The model was used to calculate the electrostatic potential surfaces around rK9 and compared with those of rK2. The complementarity of these potentials to those of aprotinin has been analysed to obtain a provisional explanation as to why rK9 is not inhibited by aprotinin (Moreau et al., 1992). The electrostatic potentials showed two positive residues in rK9 that are not present in rK2 or any other kallikreins except human hK3 (PSA). They could prevent the binding of aprotinin to rK9. A single mutation in aprotinin has been identified that might allow the binding of aprotinin to rK9.
The second step consisted of aligning the rK9 sequence with rK2, porcine kallikrein and bovine trypsin. The sequence of rK9 (Brady et al., 1989) is aligned with the reference proteins used to build the rK9 model by structural homology. The SCRs were determined as described in Materials and methods and are marked [---SCR---]. The structurally variable regions in the reference proteins are indicated as loops (loop 1–5). The amino acids are numbered at the top of the sequence according to the chymotrypsinogen numbering system (Hartley, 1970). pk1 is porcine kallikrein K1, and trp is bovine trypsin.

coordinates of these molecules were taken from the Brookhaven Protein Data Bank (PDB; Bernstein et al., 1977) files 3PTN (Bode and Schwager, 1975), 2PKA (Bode et al., 1983) and 1TON (Fujinaga and James, 1987), respectively. The model was built using the strategy developed in the HOMOLOGY module of the INSIGHT II package (Biosym Technologies Inc.). The structurally conserved regions (SCRs) among the three reference proteins were identified by an iterative procedure comparing the reference proteins one pair at a time. Regions of the reference proteins with an r.m.s. deviation <0.75 Å were designated SCRs. A set of consensus SCRs was then defined as the intersection of the individual SCRs in any given region. This set of SCRs was considered to be the framework on which the model was built. Eight SCRs were found, ranging in size from 10 residues for SCR 5 to 36 residues for SCR 7 (Figure 1). The reference proteins were superimposed over the length of their SCRs to place them all in the same coordinate frame.

The second step consisted of aligning the rK9 sequence (Brady et al., 1989) with those of reference proteins in the segments denoted by the SCRs. This alignment was performed using a Biosym proprietary algorithm because summary SCRs were defined previously among the set of reference proteins. This procedure does not allow for any insertions/deletions within the SCRs, and then incorporates some structural information into the sequence alignment. Loop 2 (kallikrein loop) in rK9 has an extra seven-residue insertion compared with the pK1 kallikrein loop (Bode et al., 1983). The same is true for rK2, but this region is not defined in its electron density map (Fujinaga and James, 1987) probably because of its great flexibility. Two deletions of two residues occur in both rK9 and rK2, one at position 36–38 (loop 1) and the other at position 147A–147B (loop 3) (the chymotrypsin numbering system is used throughout this paper; Hartley, 1970).

The framework of our model was then built by assigning coordinates from the reference proteins. The reference protein used depended on the degree of sequence similarity. In fact, the coordinates of rK9 SCRs were taken from the corresponding SCRs of rK2, which has the greatest sequence similarity to rK9. When corresponding amino acids were not the same in the model and in the chosen SCR, the side chain was replaced by keeping the same conformation as in the crystal.

Loop searches in the PDB were performed for loops 1, 3, 4 and 5 (Figure 1) to model the variable regions between the SCRs. For this purpose, a Cα distance matrix was constructed for the residues on both sides of each loop and compared with a precalculated Cα distance matrix derived from high-resolution structures extracted from the PDB.

After examination of the 10 best matches, coordinates from the rK2 loops were assigned to the corresponding loops in rK9. The kallikrein loop (loop 2) was difficult to model accurately because of the lack of an equivalent structure in the reference proteins. A search through the PDB structures (Jones and Thirup, 1986) produced no loop that fitted properly into the selected region of rK9, probably because of the length of this intervening sequence. The kallikrein loop was therefore generated de novo using the random tweak method (Shenkin et al., 1987). Different conformations were generated using geometric constraints and random values for Ψ and Φ angles to produce a group of loop structures close to an energy minimum that fitted the Cα anchor groups (Ile95 and Asp98). The loop that left the active site free for the access of substrates/inhibitors was incorporated into the model as a reasonable kallikrein loop structure.

The model was then completed by assigning coordinates for the C-terminal region (residues 240–246) directly from the corresponding region in rK2, because the sequences of both proteins are the same in this region. Hydrogen atoms were then added according to standard geometries, and disulfide bonds were added according to the same pattern in rK2 (Cys22–Cys157, Cys42–Cys58, Cys136–Cys201, Cys168–Cys182 and Cys191–Cys220).

Refinement of the rK9 model using molecular mechanics

Major steric overlaps were first removed manually and the junctions between the SCRs and loop regions were repaired using the steepest descent algorithm. Refinement was begun by energy minimization of the mutated loops (i.e. those loops with substituted residues), followed by a minimization of the side chains in the SCRs. These minimization steps were carried out until the maximum derivative was <5 kcal/mol Å. The atoms of the backbone were held tethered with a force constant of 1000 kcal/mol Å² during this step. Finally, the rK9 model was minimized unconstrained until the derivative was <0.1 kcal/mol Å. Energy minimization and molecular
dynamics were performed on a Silicon Graphics 4D25 computer using Discover (Biosym Technologies Inc.) and the consistent valence forcefield (Hagler and Lifson, 1974; Hagler et al., 1974). All calculations were performed with a 10 Å nonbond cut-off and a 2×r distance-dependent dielectric constant to approximate aqueous solvent conditions (Cardin et al., 1991), because solvent molecules were not included in the above calculations.

Checking the validity of the model
Several approaches were used to evaluate the model. First, a Ramachandran plot of the main-chain Φ-Ψ dihedral angles of the minimized rK9 model (Figure 2) indicated that 99% of the nonglycine and nonproline residues were in the allowed regions. One residue was located in a disallowed region (Asn174). This residue was adjacent to residues establishing the junction between loops and SCRs and has similar angle values in the rK2 crystal.

The accuracy of the model was then assessed by the 3-D profile method, which measures the compatibility of a protein model with its sequence (Bowie et al., 1991; Lüthy et al., 1992). The 3-D profile of the rK9 model was established using the program supplied by Dr D.Eisenberg. We also checked for the absence of buried charged residues in the core of the protein that would be energetically unfavourable, and the environment of charged residues, i.e. appropriate location and solvent-accessible surface (Lee and Richards, 1971).

Exploration of the conformational space of the kallikrein loop using molecular dynamics
The conformational space of the kallikrein loop of rK9 was searched using a combination of molecular dynamics and minimization, starting from the initial geometry of this loop generated as described above. The 13 residues of the loop (Figure 1) were allowed to move, while the rest of the structure was fixed, except for the flanking residues (residues 91, 92, 93 and 96, 97, 98) which were tethered with a force constant of 100 kcal/mol/Å². The procedure used was a simple form of simulated annealing: 2 ps of dynamics at 900 K, followed by 2 ps at 300 K. Each structure was then minimized using the steepest descents algorithm until the derivative was <5 kcal/mol/Å and the conjugate gradient algorithm until the derivative was <1 kcal/mol/Å. This procedure was repeated 50 times, allowing a 200 ps total time to be explored. All calculations were performed with Discover in vacuo using a 10 Å nonbond cut-off, a 1 fs integration time and a 1×r distance-dependent dielectric constant. As solvent molecules were not included, all the amino acids were defined to be in their neutral states to avoid the artifactual dominance of electrostatic terms (Mackay et al., 1989). The total energy was plotted versus time to analyse the resulting loop conformations.

Calculation of electrostatic potentials
The interaction of aprotinin with rK2 and rK9 was examined using the putative complex of both enzymes with aprotinin. This was obtained by superimposing separate components on the aprotinin-porcine kallikrein complex (2KAI PDB file; Chen and Bode, 1983). The SCRs between rK2 or rK9 and porcine kallikrein of the aprotinin–porcine kallikrein complex were first determined within the HOMOLOGY module of INSIGHT II. The enzymes were then superimposed onto the aprotinin–kallikrein complex using the Cα atoms in the set of SCRs. The aprotinin molecule was finally superimposed on the aprotinin component of the pK1–aprotinin complex to obtain the resulting rK9–aprotinin and rK2–aprotinin complexes. Few bad contacts occurred at the complex interface, but they were not relieved by energy minimization because they did not interfere with the qualitative analysis of electrostatic potential contours of the separate molecules. The electrostatic potentials of the enzymes and aprotinin (coordinates extracted from the aprotinin–porcine kallikrein complex) were calculated by the finite difference Poisson–Boltzmann method (Klapper et al., 1986; Gilson et al., 1987). This method takes into account the dielectric constants of the interior of the protein and of the exterior (the solvent molecules). So this is a more accurate method for analysing the electrostatic potentials of proteins than simple Coulombic charge–charge interactions in vacuo. We assumed dielectric constants of 2 for the interior of the proteins and 80 for the exterior, an ionic strength in the aqueous environment of 0.15 M and a temperature of 300 K. Calculations were performed at pH 8.5. Formal charges, rather than all charges, at pH 8.5 were assigned as follows: arginine, lysine and the N-terminus, +1; glutamate, aspartate and the C-terminus, –1; and histidine, neutral. Contours of electrostatic potential were displayed at ±0.6 kcal/mol (per unit charge). Electrostatic potentials were calculated using the DELPHI module (Biosym Technologies Inc.), and contours were displayed using INSIGHT II.

Results and discussion
Building the model
The computer program HOMOLOGY was used to build the 3-D model of rK9 by homology with the structure of rK2, pK1 (porcine kallikrein) and bovine trypsin, as described in the Materials and methods. This model building procedure, using more than one reference protein, is reported to produce more accurate models (Blundell et al., 1987, 1988; Greer, 1990) because it allows both the variable regions and SCRs within a protein family to be determined. Hence, this method is more accurate for rK9 modelling than simply replacing rK2
These moderate r.m.s. shifts are in favour of a viable model, for the backbone and for all heavy atoms between the initial as described in the Materials and methods. The r.m.s. deviations relieved before minimizing the model in a stepwise manner,

There were no major atomic steric clashes in the final rK9 Energy minimization and model evaluation segment occurs in the native rK9 and rK2 molecules. The conformation of the Gly215-Pro219 segment in rK2, conformation was obtained after a 55 ps simulation (Figure 3). The interactions of the kallikrein loop with a key residue in loop 4 could dictate the overall conformation of the kallikrein loop (Ring and Cohen, 1993). The conformation of the kallikrein loop could be closed or open, depending on the hydrophilic or hydrophobic nature of this key residue in the nearby loop 4. The presence of a putative N-glycosylation site in the kallikrein loop of all rat kallikreins might also influence the structure of this region. The kallikrein loop in rK9 was generated de novo using the random tweak method (Shenkin et al., 1987) in an attempt to predict a reasonable loop structure. The loop conformation which did not hinder the active site was selected for further analysis of ligand binding.

The interaction of the kallikrein loop with a key residue in loop 4 could dictate the overall conformation of the kallikrein loop (Ring and Cohen, 1993). The conformation of the kallikrein loop could be closed or open, depending on the hydrophilic or hydrophobic nature of this key residue in the nearby loop 4. The presence of a putative N-glycosylation site in the kallikrein loop of all rat kallikreins might also influence the structure of this region. The kallikrein loop in rK9 was generated de novo using the random tweak method (Shenkin et al., 1987) in an attempt to predict a reasonable loop structure. The loop conformation which did not hinder the active site was selected for further analysis of ligand binding.
The closed conformation obtained at 55 ps supports this exclusion of water molecules from the active site (Weber, Tilbeurgh et al., 1994). This function of the kallikrein loop is similar to that of the flap of the human immunodeficiency virus protease, which contributes to inhibitor binding and the exclusion of water molecules from the active site (Brillard-Bourdet et al., 1995) that loop 4 is the main determinant of the extended enzymatic activity of rat kallikreins.

The regions where there are notable structural differences between rK9 and the reference proteins are located in surface loops surrounding the active site (Figure 4). These loops also correspond to the five most variable regions in a multiple alignment of all tissue kallikreins described to date (Wines et al., 1991). Kinetic studies have shown (Brillard-Bourdet et al., 1995) that loop 4 is the main determinant of the extended interaction site of rat kallikreins, with residues of the substrate hypothesis, as the 3-D scores increase towards positive values for this peculiar region (Figure 5). Comparison with the equivalent loop in the PSA model (PDB file 1PFA) also suggested that this loop is very flexible. Although both loops are the same length, the minimal r.m.s. deviation is as high as 5.1 Å after superimposing the Cα atoms (Figure 4).

All rat tissue kallikreins, except for rK1 and rK2, may be cleaved at a peptide bond involving an arginyl residue in the kallikrein loop. This generates an N-terminal light chain and a C-terminal heavy chain. Similar cleavages occur in human and pig kallikreins, but the biological relevance of this cleavage is not clear because the cleaved enzymes do not lose their enzymatic activity in vitro. The only two kallikreins that have a known biological function, rK1 and rK2, are single-chain enzymes. Therefore the kallikrein loop could modulate the enzymatic activity of rat kallikreins in vivo.

Comparison of the model with the template proteins

The rK9 model and the pK1 structure are superimposed in Figure 4. The r.m.s. deviation between the two structures is 0.90 Å after superimposing 212 Cα atoms in the SCRs, and 0.96 Å after superimposing backbone atoms. These values are in accordance with the expected value of 1.06 Å for two proteins having 48% sequence identity (Chothia and Lesk, 1986). The r.m.s. deviations for 222 Cα atoms and 888 backbone atoms between rK9 and rK2 are 0.35 and 0.42 Å, respectively, close to the expected value of 0.54 Å. Thus, the amino acid sequence of rK9 is consistent with a folded conformation similar to those of rK2 and pK1.
differs from that of the other serine proteinases in the Gly215-Pro219 segment. This segment normally defines a part of the S1 subsite, but allows the opening of a new large pocket between this Gly215-Pro219 segment and the Lys224-Tyr228 segment. This new subsite is still in the immediate vicinity of the catalytic triad and may form the S1 subsite of rK9. The Cα trace of rK9 is shown as a thick line, and that of pK1 as a thin line.

Fig. 6. The S1 binding pocket of rK9 and pK1. The conformation of rK9 differs from that of the other serine proteinases in the Gly215-Pro219 segment. This segment normally defines a part of the S1 pocket. It is shifted in rK9 and in rK2. This shift leads to a reduction in the size of the S1 subsite, but allows the opening of a new large pocket between this Gly215-Pro219 segment and the Lys224-Tyr228 segment. This new subsite is still in the immediate vicinity of the catalytic triad and may form the S1 subsite of rK9. The Cα trace of rK9 is shown as a thick line, and that of pK1 as a thin line.

beyond position P4 [the Pn and Sn nomenclature for substrate and enzyme subsites, respectively, was introduced by Schechter and Berger (1967)]. Loops 1 and 3, which are two residues shorter in rK9 than in pK1, are also likely to play a crucial role in determining the extended recognition selectivity for substrate P′ residues. The kallikrein loop (loop 2) is also believed to modulate the specificity for substrate residues N-terminal to the scissile bond. The Asp189 at the bottom of the specificity pocket is thought to confer to trypsin-like enzymes their P1, Arg specificity (Bode et al., 1984). However, recent studies indicate that distal portions of the enzyme are also crucial for determining the trypsin- or chymotrypsin-like specificity of a serine proteinase (Perona et al., 1995). This agrees with the observation that rK9 and rK2 both have an aspartyl residue at position 189 and have trypsin- and chymotrypsin-like specificity (Moreau et al., 1992).

Structure of the active site and consequences for ligand binding

rK2 and rK9 are unique among rat kallikreins in that they have both trypsin- and chymotrypsin-like specificities, cleaving substrates with either arginyl or aromatic residues at P1. The other tissue kallikreins with chymotrypsin-like activity is human PSA (hK3; Dubé, 1994). A recent analysis of a PSA structural model has shown that this specificity is determined by the shape of the S1 pocket, which resembles that of chymotrypsin, and by the lack of Asp189 at the bottom of this pocket, which prevents arginyl side chains from binding in this pocket (Villoutreix et al., 1994). Asp189 is present in both rK9 and rK2, but the conformation of their S1 pockets differs significantly from those of other serine proteinases. The Gly215-Pro219 segment, which normally forms the opening of the S1 pocket in homologous serine proteinases, is shifted in rK9, leading to a dramatic reduction in this pocket between this segment and the Thr190-Ser195 segment (Figure 6). The modified S1 pocket is no longer able to accommodate arginyl side chains or large aromatic groups of the substrates. These are accommodated by the S1 pocket of pK1 complexed to benzamidine (Bode et al., 1983). However, this shift of the Gly215-Pro219 fragment generates a new pocket in rK9 between this segment and the Lys224-Tyr228 segment, as in rK2 (Fujinaga and James, 1987). This new S1 pocket remains in the immediate vicinity of the catalytic triad and is large enough to accommodate large aromatic groups as well as an arginine side chain.

We attempted to verify this hypothesis by identifying ligands able to fill this pocket using the ligand design software LUDI (Biosym Technologies Inc.). LUDI selects fragments from a library that form specific interactions (hydrophobic, electrostatic and hydrogen bonds) with residues at a given interaction site. The search was performed in a sphere of 12 Å diameter centred on a pseudo-atom placed at about the middle of the pocket to be searched. Surprisingly, benzamidine was the ligand with the best score. Its accommodation within this newly formed S1 pocket is such that the amidine group interacts electrostatically with Asp189 at the bottom of the pocket. Seven hydrogen bonds are formed between benzamidine and residues 187, 188, 215 and 218 of rK9. The positioning of the benzamidine molecule clearly prevents substrates from accessing the active site (Figure 7), and thus accounts for the inhibition of rK9 by benzamidine (C.Wang et al., 1992).

The side chains of arginine, phenylalanine and tyrosine also interact within this pocket with good score values. The same search within the equivalent pocket in rK2 gave similar results. This is in agreement with the dual enzymatic specificity of
Molecular model of rat kallikrein rK9

Fig. 8. Electrostatic surface potentials of rat kallikreins rK2 and rK9, and aprotinin. The contours of electrostatic potentials around the enzymes are shown at levels of -0.6 (red) and +0.6 kcal/mol (dark blue). The contours around aprotinin are shown at levels of -0.6 (magenta) and +0.6 kcal/mol (light blue). The enzyme molecules are rotated 90° clockwise around the y-axis, compared with the orientation shown in Figure 4, and are coloured green for rK2 and yellow for rK9. As expected for rK2 (net charge -3), the enzyme is covered with large regions of negative potential and is surrounded by small regions of positive potential. The electrostatic potential contours around rK9 (net charge 0) are similar to those of rK2, except for a region of positive potential near the entrance of the active site which is considerably larger in rK9 than in rK2. This large positively charged area is caused by Lys61 and Arg65 in rK9. The interaction of aprotinin with both rK2 and rK9 was studied by building the putative rK2-aprotinin and rK9-aprotinin complexes by superimposing the individual molecules onto the equivalent components of the crystal pKl-aprotinin complex (Chen and Bode, 1983). The aprotinin molecule in the putative rK2-aprotinin and rK9-aprotinin complexes has been translated along the ;y-axis for clarity. Aprotinin (net charge +6) interacts with the negative core of the senne proteinase via its positively charged inhibitory site. The negative core of rK2 is complementary to the positive potential of aprotinin at the complex interface, thus contributing to the efficient binding of the positively charged aprotinin. In the rK9-aprotinin complex, the positively charged residues Lys61 and Arg65 of the enzyme induce an unfavourable positive-positive overlap (see Figure 9), which may explain why this inhibitor does not inhibit rK9. The electrostatic potentials were calculated for the isolated molecules using the Poisson—Boltzmann method implemented in DELPHI (Biosym Technologies Inc.), assuming dielectric constants of 2 for the interior of the proteins and 80 for the exterior, an ionic strength of 0.15 M and a temperature of 300 K. CPK representations of Lys61 and Arg65 are shown in yellow.

both rK2 and rK9 (Moreau et al., 1992). However, the binding mode of these side chains is such that the conformation of the backbone at the scissile bond would not be optimal for this bond to be cleaved. This could explain the relatively low catalytic activity of both rK2 and rK9 towards peptide substrates compared with the other tissue kallikreins (Moreau et al., 1992).

A similar movement of the binding loop (loop 5) has been reported for human Factor D, a serine protease of the complement system (Narayana et al., 1994). This shift would be responsible for the peculiar reactivity and substrate specificity of Factor D. The authors also suggest that the peculiar conformation of the binding loop could explain the low reactivity of this enzyme for peptide substrates, which parallels the relationship between rK2 and rK9, and other rat tissue kallikreins (Gauthier et al., 1992). The model may also explain why rK9 cannot accommodate bulky, hydrophobic residues at position P2 of the substrate, as do most other kallikreins.

The P2 specificity of tissue kallikreins is thought to be the result of the presence of a wedge-shaped hydrophobic cavity between Tyr99 and Trp215 which traps hydrophobic P2 residues (Chen and Bode, 1983; J.Wang et al., 1992). The residues Tyr99 and Trp215 are replaced by a histidyl and a glyeiy residue, respectively, in rK9 and rK2, so that the substrate hydrophobic P2 residue is no longer required. Instead, rK9 and rK2 preferentially accommodate a Pro at that position, probably to orientate the scissile bond in a suitable conformation for subsequent cleavage. This may also account for the rapid hydrolysis of the Z-Val-Lys-Lys-Arg-AFC substrate by rK9 and rK2, which has a hydrophilic residue in P2 (Shori et al., 1992).

The model also shows that the cluster of hydrophobic residues formed by Ala32, Ile35, Phe41, Leu67, Leu73, Val74, Trp141, Met149 and Leu151 in both rK9 and rK2 is an essential determinant for the preferential binding of hydrophobic residues at positions P2′-P4′ (Moreau et al., 1992), as first suggested (Fujinaga and James, 1987) to explain the cleavage of angiotensinogen by rK2.

Probing the lack of inhibition of rK9 by aprotinin by comparing electrostatic potentials

Electrostatic forces are important for molecule recognition. They contribute significantly to the specificity of protein—protein (Perry et al., 1989) and protein—ligand (Klapper et al., 1986; Bajorath et al., 1991) interactions, which often involve complementary charge distributions on the binding interfaces, as in the thrombin—hirudin complex (Karshikov et al., 1992).
Previously we showed that aprotinin binds rK2 with an apparent inhibition constant of 76 μM, whereas rK9, despite its structural similarity with rK2, is not inhibited by this inhibitor (Moreau et al., 1992). The interaction of aprotinin with both enzymes was investigated further by calculating the electrostatic potentials of each partner in aqueous solution. rK9–aprotinin and rK2–aprotinin complexes were built, starting from the crystal structure of the kK1–aprotinin complex (Chen and Bode, 1983). The enzymes were first superimposed on the kK1 component of the kK1–aprotinin complex. The aprotinin molecule was then superimposed on the aprotinin component of the kK1–aprotinin complex to obtain the resulting rK9–aprotinin and rK2–aprotinin complexes. The contours of electrostatic potentials (at −0.6 and +0.6 kcal/mol) for rK9 and rK2 are shown in Figure 8. rK2 carries a net charge of −3, whereas the overall net charge of rK9 is neutral. However, the charged residues in both rK2 and rK9 are distributed in a similar way. Both enzymes comprise mostly a negative core, surrounded by a few spots of positive potential (Figure 8). At variance with rK2, rK9 has a large positive area at the surface of the molecule because of the presence of Lys and Arg at positions 61 and 65 respectively. The electrostatic potentials of aprotinin were also determined so as to find a structural basis for the lack of aprotinin binding to rK9. Aprotinin, which is essentially positive (net charge +6), interacts with serine proteinases via a large positively charged area that complements the negatively active site of the proteinases (Figure 8). The interaction between aprotinin and both proteinases can be seen in detail in Figure 9, which shows a slice through the contours of electrostatic potential superimposed on the structures of the inhibitor–protease complexes. The potential of the inhibitor is highly complementary to the potential of rK2 at the binding region, which agrees with the tight binding of aprotinin to rK2. The same result was obtained for the pK1–aprotinin crystal complex after calculation of the electrostatic potentials for the separate molecules of the complex (data not shown). The contours of electrostatic potential of the rK9–aprotinin complex reveal a different situation. The large positive region found only in rK9 appears to be close to the putative aprotinin binding site in rK9, so that it overlaps the positive potential of aprotinin. This large and unfavourable positive–positive overlap...
overlap may partly explain the absence of aprotinin binding to rK9. This is supported by the fact that rK9 is the only rat tissue kallikrein not inhibited by aprotinin, and is also the only enzyme of the family to have positive residues at positions 61 and 65 (Figure 10). In agreement with this observation, human PSA also has two positively charged residues at positions 60 and 63 (Figure 10) and interacts very weakly with aprotinin (Christensson et al., 1990). The electrostatic potential contours displayed for a PSA model (Villoutreix et al., 1994) clearly indicate the presence of the same over-lapping positive region as in rK9. pK1 has a lysine at position 59, but the two negatively charged residues (Glu and Asp) at positions 61 and 65 generate a large area of negative potential. The electrostatic surface potential of pK1 (Villoutreix et al., 1994) is thus highly complementary to that of aprotinin, which accounts for the tight binding of aprotinin to pK1. On the other hand, Struthers et al. (1991) showed that the interaction of the amyloid precursor protease inhibitor domain with trypsin is modulated by repulsive electrostatic forces. It is therefore very likely that long-range attractive/repulsive electrostatic forces are an important factor in selecting the protease inhibitors that bind to serine proteinases.

Molecular modelling studies can be used to check the importance of electrostatic forces for inhibitor binding. A single mutation in aprotinin would allow the binding of aprotinin to rK9. The Lys46 residue in aprotinin is involved in the positive–positive overlap with the Lys61 and Arg65 residues of rK9. The electrostatic field of a mutant aprotinin in which Lys46 is replaced by a glutamate residue indicates that the corresponding region of positive potential in the inhibitor would be replaced by an area of negative potential (Figure 9). The Lys/Glu46 aprotinin can thus be expected to inhibit rK9.

Acknowledgements

We thank Drs Jean-Dominique Guitton (Rhône Poulenc Rorer) and Jair Chagas (Escola Paulista de Medicina, São Paulo, Brazil) for valuable discussions. We also thank Drs Dave Kiszon and Justin Taylor of Biosym Technologies for technical assistance in electrostatic calculations and computer modelling, and Dr David Eisenberg for supplying the Fortran source code of the 3-D profile suite program. The English text was edited by Dr Owen Parkes. The coordinates of the rK9 model will be deposited at the PDB.

References
