A structural model for the prostate disease marker, human prostate-specific antigen

BRUNO O. VILLOUTREIX,¹ ELIZABETH D. GETZOFF,² AND JOHN H. GRIFFIN¹

¹ Departments of Molecular and Experimental Medicine and Vascular Biology, The Scripps Research Institute, La Jolla, California 92037

² Department of Molecular Biology, The Scripps Research Institute, La Jolla, California 92037

(RECEIVED July 14, 1994; ACCEPTED September 2, 1994)

Abstract

Prostate-specific antigen (PSA) provides an excellent serum marker for prostate cancer, the most frequent form of cancer in American males. PSA is a 237-residue protease based on sequence homology to kallikrein-like enzymes. To predict the 3-dimensional structure of PSA, homology modeling studies were performed based on sequence and structural alignments with tonin, pancreatic kallikrein, chymotrypsin, and trypsin. The structurally conserved regions of the 4 reference X-ray proteins provided the core structure of PSA, whereas the loop structures were modeled on the loops of tonin and kallikrein. The unique "kallikrein loop" insert, between Ser 95b and Pro 95k of kallikrein, was constructed using molecular mechanics, dynamics, and electrostatics calculations. In the resulting PSA structure, the catalytic triad, involving residues His 57, Asp 102, and Ser 195, and hydrophobic and electrostatic interactions typical of serine proteases were extremely well conserved. Similarly, the 5 disulfide bonds of kallikrein were also conserved in PSA. These results, together with the fact that no major steric clashes arose during the modeling process, provide strong evidence for the validity of the PSA model. Calculation of the electrostatic potential contours of kallikrein and PSA was carried out using the finite difference Poisson-Boltzmann method. The calculations revealed matching areas of negative potential near the catalytic triad, but differences in the positive potential surrounding the active site. The PSA glycosylation site, Asn 61, is fully accessible to the solvent and is enclosed in a positive region of the isopotential map. The bottom of the substrate specificity pocket, residue S1, is a serine (Ser 189) as in chymotrypsin, rather than aspartate (Asp 189) as in tonin, kallikrein, and trypsin. This fact, plus other features of the S1 binding-pocket region, suggest that PSA would prefer substrates with hydrophobic residues at the P1 position. The location of a potential zinc ion binding site involving the side chain of histidines 91, 101, and 233 is also suggested. This PSA model should facilitate the understanding and prediction of structural and functional properties of this important cancer marker.

Keywords: electrostatic potentials; homology modeling; prostate-specific antigen; PSA; serine protease

Prostate-specific antigen (PSA) initially was discovered in semen by Hara et al. (1971), who named it γ -seminoprotein, and further studied by Li and Beling (1973) and Sensabaugh (1978). It has received acceptance as a useful forensic marker of rape (Grave et al., 1985). However, the widest clinical interest in PSA is centered in its remarkable value as a serum marker for carcinoma of the prostate (Wang et al., 1979; Stamey, 1989; Brawer, 1991; Gittes, 1991; Oesterling, 1991). PSA is produced uniquely by the prostate and is made by normal, hyperplastic, and malignant prostatic tissue. In normal men under 40 years of age, serum PSA is <2 ng/mL. Circulating PSA levels can be significantly elevated in benign and malignant prostate diseases. Some

have accorded it to be the most useful marker in all of tumor biology.

PSA lyses the coagulum formed upon ejaculation when prostatic fluid is mixed with seminal vesicle fluids (Lilja, 1985). This gelation mainly involves coagulation of semenogelins I and II, proteins derived from seminal vesicle secretions, which are subsequently cleaved by PSA at tyrosine and leucine residues during the lytic reaction (Lilja et al., 1989, 1992). PSA stimulates the growth of normal and cancerous prostate cells in culture (Cohen et al., 1992), suggesting that one possible mode of PSA action on growth may involve the downregulation of an insulin-like growth factor binding protein (IGFBP) by proteolysis of IGFBP-3 or IGFBP-4 by PSA. PSA may be, in effect, an autocrine or paracrine growth-promoting factor for prostate cells.

PSA is likely synthesized as a preproenzyme that is very similar to tissue preprokallikrein, except that it has chymotrypsin-

Reprint requests to: John H. Griffin, The Scripps Research Institute, Molecular and Experimental Medicine (SBR5), 10666 North Torrey Pines Road, La Jolla, California 92037; e-mail: bruno@scripps.edu.

like (Akiyama et al., 1987; Christensson et al., 1990) rather than trypsin-like specificity based on protein and cDNA sequencing studies (Watt et al., 1986; Lundwall & Lilja, 1987; Schaller et al., 1987; Riegman et al., 1988). PSA is a glycoprotein of 33,000 MW and the proenzyme (zymogen) (244 residues) likely has an activation peptide of 7 amino acids, APLILSR, that is released by as yet unknown protease(s) to generate enzymatically active PSA (237 residues).

In the past few years, homology modeling carried out on various proteins has demonstrated that functionally significant regions of related proteins can be predicted with reasonably high accuracy. This method is a valuable tool, particularly when several homologous experimental structures are known (Furie et al., 1981; Greer, 1981, 1990; Fisher et al., 1994). PSA, like kallikrein, is a member of the serine protease family of enzymes whose prototypes include chymotrypsin and trypsin, which have a characteristic catalytic triad of serine, histidine, and aspartic acid residues in the enzyme-active site. In this article, we homology model the 3-dimensional structure of PSA using known X-ray crystallographic structures. The numbering of PSA was modified to agree with the standard chymotrypsinogen numbering system used for serine proteases (Fujinaga & James, 1987). The PSA sequence contains an insertion of 7 residues relative to the actual X-ray structure of kallikrein and 8 relative to the X-ray structure of tonin, in the region of residue 95 known as the "kallikrein loop" (Bode et al., 1983; Fujinaga & James, 1987). Tonin and kallikrein also have insertions in this region of the protein when compared to chymotrypsin or trypsin, but X-ray data for this particular loop are not available for these enzymes due to crystal disorder in the case of tonin and cleavage in kallikrein. The loop structure of PSA was therefore constructed using molecular mechanics and dynamics simulations in conjunction with the solvation energy. The resulting model was used to calculate the electrostatic potential surfaces around the PSA enzyme, which were then compared to that of kallikrein. Finally, the structure of bovine trypsin inhibitor-kallikrein complex (Chen & Bode, 1983) was used to dock 6 residues of a putative substrate in the active site of PSA.

Results and discussion

Protein modeling

Overall structure description of serine protease enzymes

The structure of the serine protease enzymes consists of 2 6-stranded antiparallel β -barrels with the active site located in a cleft between these 2 barrels. Serine proteases have 3 α -helical segments, the first between the third and fourth strands in barrel 1 (residues 55–59), the second in the connecting segment between the 2 barrels (around residues 165–175), and the third at the C-terminus. Known X-ray structures of serine proteases contain 4–6 disulfide bridges. The main differences between known X-ray structures of serine proteases are mainly located in several loop regions around the active site, and these structural differences are of importance for their enzymatic specificities.

Structural and sequence alignment

The structural alignment (Fig. 1) of the 4 reference proteins (chymotrypsin, trypsin, tonin, kallikrein) obtained was similar

to that of Greer (1990). The C α RMS deviations among the average structurally conserved regions (SCRs) of the 4 different protein structures are indicated in Table 1. The sequence of the PSA enzyme was then aligned with that of the reference set (Fig. 1). The final alignment (Fig. 1) introduced an insertion for the PSA sequence relative to the other members of this family, except for tonin, in the so-called "kallikrein loop" at residue 95. Among the 4 reference proteins, only tonin has a similar insertion in this part of the molecule, but no structural data are available for the kallikrein loop region of tonin (Fujinaga & James, 1987) or kallikrein (Bode et al., 1983). PSA also has a deletion of 2 residues relative to kallikrein in the loop region near residue 145.

The sequences of these proteins of known structure show a degree of identity with the amino acid sequence of PSA varying from 35% (chymotrypsin) to 55% (kallikrein). The most striking observations that emerge from these comparisons are that the amino acid residues known to constitute the active site (catalytic triad) of the serine proteases from the crystallographic data and site-directed mutagenesis experiments are highly conserved and that important residues involved in stabilization of the structure are also conserved, including disulfide bridges, hydrophobic interactions and charge-charge interactions, as well as the residues immediately flanking the active site and those within the SCRs.

Model building

The computer programs Homology and InsightII (Biosym Technologies, Inc.) were used for constructing a 3-dimensional (3D) model of PSA from the structures of kallikrein and tonin as detailed in the Methods. The most difficult part of the model building was the construction of the 8-amino acid residue insertion, the kallikrein loop of PSA, between residue Ser 95b and Pro 95k at the border of 2 SCRs and the active site (Fig. 1). A full investigation of the potential energy surface of this region would have been computationally prohibitive, especially if solvent molecules were added. The loop-building method of Jones and Thirup (1986), which attempts to fit a structure from a non-homologous protein with the required number of residues into the model, was used first, but no good fits were found, probably due to the long length of this loop. We then used the "random tweak" algorithm (Shenkin et al., 1987), which generates a loop de novo by randomly searching conformational space for a suitable backbone conformation (see Methods). A suitable loop conformation was defined as one that left the active site open and that did not create unfavorable steric clashes with the rest of the structure. The resulting structure was then submitted to conformational search via molecular dynamics simulations.

The region of the 2 residues deleted in the PSA sequence relative to kallikrein in the loop region at residue 145 was modeled using the structure of tonin, which has the same length (Fig. 1). The rest of the model building was straightforward because no severe clashes occurred.

Energy minimization

Energy minimization was used to refine the PSA model by relieving conformational strain. The crystallographic structure of kallikrein was also energy minimized as a control for evaluating the PSA model. Initially, the major component of the potential energy for both structures arose from unfavorable van

Fig. 1. Sequence and structural alignment of PSA with trypsin (TLD), chymotrypsin (CHA), kallikrein (PKA), and tonin (TON). The sequences of the reference proteins were aligned according to the SCRs (Homology program) and are delineated by black boxes. The sequence of PSA was then aligned by eye with the rest of the reference set, avoiding the introduction of gaps and insertions within the SCRs. Black asterisks indicate the catalytic triad (His 57, Asp 102, Ser 195) and Ser 189 at the bottom of the S1 specificity pocket. The @ symbol above the sequences depicts solvent-accessible side chains of kallikrein, whereas the absence of this symbol indicates buried side chains (see Methods). This solvent accessibility pattern is well conserved in the 4 X-ray structures as well as in the PSA model. Beneath the sequences, the black arrows represent the β -strands in kallikrein and the solid gray line indicates the α -helices. The thin dotted line in the kallikrein loop region indicates the residues present in the sequence of tonin but absent from the X-ray data. The short black vertical line before alanine 149 in the sequence of chymotrypsin represents the cleavage site in the autolysis loop. Chymotrypsin numbering is maintained throughout this sequence alignment. The kallikrein loop sequence insertion of PSA is denoted as residues 95b-95k, i.e., Ser 95b-Pro 95k (SLLKNRFLRP).

der Waals contacts that were for the most part removed after a few cycles of steepest descents minimization. The final PSA model and the minimized kallikrein X-ray structure had similar values for the covalent and noncovalent energy terms and total energies. As previously cautioned by Novotny et al. (1988), even an incorrect structure can converge to a comparatively low-energy minimum, and therefore such energy terms must be used with caution to judge the validity of a model structure. RMS deviation and localization of charged and hydrophobic residues

(see below, Evaluation of the model) provide better indicators to evaluate a model structure (Blundell et al., 1987). The RMS deviation for the backbone and for all heavy atoms between the initial and minimized PSA models were 0.89 Å and 1.16 Å, respectively. This is comparable to the RMS deviations between the kallikrein X-ray structure and its minimized structure, which were 0.80 Å and 1.02 Å, respectively. When only residues that did not belong to the surface loop regions were compared, the RMS shifts for the backbone atoms of kallikrein and PSA were 0.67 Å and 0.71 Å, respectively. These low RMS shifts support the validity of the model structure. As expected, the major movement of the backbone occurred in the surface loop regions because bulk solvent, counterions, and crystal contacts were ignored in the calculations. A Ramachandran plot of the minimized PSA model (Fig. 2) showed that most of the residues had acceptable backbone conformations.

Refinement of the kallikrein loop of PSA

Nonhomologous loop-building presents a major problem in homology modeling, and the uncertainty of the model is greatest in these regions. Nevertheless, several techniques have been shown to predict reasonable loop structures. Among these methods is the well-known geometric search through the high-resolution Protein Data Bank (Bernstein et al., 1977) developed by Jones

Table 1. RMS deviations between the C α positions in the average SCRs as evaluated for pairwise comparisons of the 4 reference protein structures^a

RMS	CHA	TON	TLD	PKA
CHA	0	0.78	0.67	0.79
TON		0	0.65	0.47
TLD			0	0.63
PKA				0

^a CHA refers to chymotrypsin, TON to tonin, PKA to kallikrein, and TLD to trypsin.

Fig. 2. Ramachandran plot of the minimized PSA model structure. The phi and psi angles for the non-glycine residues are represented as crosses and the glycine residues are shown as squares.

and Thirup (1986). This method generally gives good results for short loop sequences and will become more accurate in the future as the number of different representative structures available in the Protein Data Bank increases. Other methods using geometric and energetic criteria have also been used successfully (Fine et al., 1986; Summers & Karplus, 1990). Here we used conformational search and energetic criteria to select a plausible structure for the kallikrein loop of PSA. An initial loop was generated using the "random tweak" algorithm and introduced into the model as a starting point for the search. A high-temperature molecular dynamics simulation sampled conformational space. The dynamics method chosen gave rise to 50 different conformations, with an RMS deviation for the C α atoms of up to 4 Å when compared to the initial loop structure. The differences in the conformations were maximal at the beginning of the simulation, whereas at the end the structures seemed to oscillate near the active-site entrance without significant structural change. Two different starting conformations obtained from the "random tweak" algorithm search for this loop region were submitted to a 60-ps molecular dynamics simulation at room temperature. The results of these simulations also indicated that the loop moved toward the active site and covered it (data not shown). The possibility that the kallikrein loop might be bulkier than the actual segment of the kallikrein X-ray structure was suggested by Bode et al. (1983), in agreement with our calculations. However, it is important to note that this calculated conformation can be artifactual from the simulation because nonbonded energy terms usually tend to dominate the final conformational energy search and thus lead, for example, to a conformation with immobilized charged residues with reduced solvent accessibility. This transfer of charge from the aqueous solvent into a protein interior is energetically unfavorable (Rashin & Honig, 1984). Thus, in order to select a more probable loop structure, we examined the potential and solvation energies (Fig. 3). Consideration of solvation energy calculations together with the

results of the potential energy should help to select most reasonable loop structures (Fine et al., 1986). The lowest potential energy (at 29 ps in Fig. 3A) obtained during the dynamics simulation occurred when the loop fully covered the active-site region, whereas the structure with the lowest solvation energy (at 8 ps in Fig. 3A) left the active site accessible to a small substrate (Fig. 4B). The inverse protein folding calculations, as developed by D. Eisenberg and co-workers (Bowie et al., 1991), showed that both the lowest potential energy and the lowest solvation energy PSA structures (Fig. 4B) had mainly positive scores and therefore may represent the protein in "closed" and "open" conformations, respectively. The lower solvation energy structure was assumed to be more appropriate for structural analysis of the active site and for electrostatic potential calculations. However, the kallikrein loop of PSA is predicted to be flexible and it can be hypothesized that this loop could open up even more due to binding of macromolecular substrates or cofactors.

Overall comparison of the PSA model with X-ray structures of serine proteases

The N-terminal loop of the PSA model starts with residue Ile 16 and, like several other serine proteases, has its N-terminal amino group forming a buried ion pair with the carboxylate of Asp 194. This ion pair likely contributes to the stabilization of this region. The short helical segment containing the active-site His 57 is stabilized by the conserved disulfide bridge between the nearby residues Cys 42 and Cys 58. This part of the structure is very stable, and no shift was detected after energy minimization in the kallikrein and PSA model structures. The main difference in this first β -barrel domain between PSA and the 4 reference X-ray structures was the size of the last loop, residues 95–103, including the active-site Asp 102. This region in PSA may have several conformations and may contribute to or interfere with binding of macromolecular substrates or cofactors which reciprocally may alter the active site and activity of PSA. Residues 109–130 constitute a connecting segment that links the

Fig. 3. Refinement of the kallikrein loop of PSA. **A:** Plot of the potential energy (Epot) versus time for the high-temperature dynamics simulation (1,000 K) of the PSA kallikrein loop (residues 95–99). Snapshots of the 50-ps simulation were taken at 1-ps intervals, and each structure was annealed to 300 K and energy minimized. The lowest potential energy was obtained at 29 ps. The corresponding loop structure fully covered the active-site entrance. **B:** Plot of the corresponding solvation energy (Esolv) resulting from the dynamics simulation. The lowest solvation energy was obtained at 8 ps. The resulting loop structure leaves the active site "open" and able to bind a small oligopeptide substrate. This structure was used for further analysis of the PSA model.

Fig. 4. Molecular dynamics simulation of the kallikrein loop of PSA and verification of the structures by the inverse protein-folding approach. **A:** Results of the dynamics simulation of the PSA loop from residues 95 to 99. Solid ribbon representation of PSA. The initial loop structure is shown in white. The catalytic triad residues are shown in purple (Asp 102, His 57, and Ser 195, from left to right). The conformation having the lowest potential energy structure for the kallikrein loop obtained during the dynamic simulation is shown in blue. This structure fully covers the active site. The structure with the lowest solvation energy is shown in red. This "open" conformation of the kallikrein loop of PSA allows the docking of a small peptide substrate in the active site. This open conformation was considered more suitable for subsequent structural analyses. **B:** Comparison of the calculated 3D profiles using a 10-residue window for the PSA model. The *x*-axis depicts the residue number and the *y*-axis depicts the 3D score. The horizontal black line represents the 0 value of the score *S*, whereas the 2 vertical black lines enclose the loop at residue 95. The lowest conformational energy (gray curve) and the lowest solvation energy (black curve) correspond to the dynamics simulation superimposed except for the region of the loop at residue 95. This graph shows that the PSA models have an acceptable overall conformation (positive score).

2 β -barrels together. The first loop in the second β -barrel, corresponding to the "autolysis" loop of chymotrypsin, comprised of residues 147–152, is 2 residues shorter in PSA compared to kallikrein, but is the same length as tonin, which was used to model this loop. The loop containing the active-site Ser 195 is conserved in the model and lies in a region in which no significant shift occurred after energy minimization. Serine 189 is at the bottom of the S1 specificity pocket. Of the reference set of

proteins, only chymotrypsin has a serine residue at this position, whereas others have aspartic acid. This region is important in determining substrate specificity and, in the case of PSA, is predicted to cause a preference for cleavage at hydrophobic or aromatic residues (Bode & Huber, 1991; Lilja et al., 1992). The loop containing residues 214–228 has the same length in kallikrein and PSA. The C-terminus of PSA, Pro 237, is stabilized by interaction of its carboxyl group with the guanidine group of Arg 107.

Evaluation of the model

The model was evaluated, first, by analyzing the RMS deviation between the initial and the minimized model structures; second, by considering the appropriateness of the location of the charged, conserved, variable, and hydrophobic residues; and third, by using the inverse protein folding verification approach (Bowie et al., 1991).

RMS shift and location of charged residues

The RMS deviation (reported above under Energy minimization) between the initial and energy-minimized structure was low and was consistent with a reasonable model structure. Another indicator of the correctness of this model was the environment of the charged residues. The location and solvent accessibility (Lee & Richards, 1971) of Glu, Asp, Lys, and Arg residues were determined in the PSA model structure. Of the 10 arginine residues, 8 are exposed to the solvent. Among the less accessible arginine groups, Arg 70 interacts with Glu 77 and Arg 107 forms an ion pair with the carboxyl group of the C-terminal residue. Among the 12 lysines, only Lys 230 has limited solvent accessibility, and it forms an ion pair with Glu 128. Of the 11 glutamate residues of the PSA model, 10 are accessible to the surface. Glu 77 is less accessible but interacts with Lys 24 and Arg 70. Ten of 12 aspartate residues are very accessible to solvent. Among those less exposed, the active-site residue Asp 102 interacts with His 57, and Asp 194 with the α -amino group of the N-terminal residue.

Distribution of conserved residues and conservative substitutions

The amino acids present in the SCRs (Fig. 1) are generally conserved or display conservative substitution. As suggested by Greer (1985) and Struthers et al. (1991), a strong preponderance of conserved residues in the interior of a modeled protein structure gives a high level of confidence in the validity of the structure. This is the case for the final model structure of PSA and attests to the reasonableness and the accuracy of the model. The 5 disulfide bridges critical to the structure of kallikrein were conserved in PSA. The disulfide bonds involved are Cys 22–Cys 157, Cys 42–Cys 58, Cys 136–Cys 201, Cys 168–Cys 182, and Cys 191–Cys 220. As an example of conservative mutation, the patch of hydrophobic interactions involving Ala 32, Leu 73, and Trp 141 in kallikrein were conserved in PSA and involve Leu 32, Leu 73, and Trp 141. Other hydrophobic interactions involving Val 52 and Val 65 in kallikrein were conserved in PSA between Val 52 and Ile 65. Similar conservation was found in the hydrophobic contact between residue Leu 163 and Leu 185 of kallikrein, which involved Ile 163 and Trp 185 in PSA. In addition, several charge–charge interactions were also conserved in PSA compared to kallikrein. These interactions conserved in kalli-

Fig. 5. Electrostatic surface potential of kallikrein. **A:** Kallikrein is shown in a green ribbon representation. The catalytic triad residues and the aspartate 189 present at the bottom of the S1 specificity pocket are shown in red. The loop at residue 95 is shown in blue and the loop containing residue 145 in purple. **B:** In a similar orientation, isopotential contours at a level of -1 kcal/mol (in red) and $+1$ kcal/mol (in blue) are seen. Kallikrein has a net charge of -13.5 and is largely covered by negative contours.

krein and PSA were Asp 116–Lys 119, Lys 175–Asp 98, Lys 24–Glu 77–Arg 70, and Lys 230–Glu 128. The 3D profile obtained by the inverse protein folding method (Fig. 4B) showed that the overall conformation of the PSA model is acceptable.

Together with the fact that no major steric clashes appeared during the construction of the model structure, the above results provide strong evidence for the validity of the model and therefore this structure was judged appropriate for further detailed analysis.

Electrostatic potential contours

Electrostatic effects are important factors in determining not only the conformation and stability of proteins in solution, but

also the binding of ligands (Sharp & Honig, 1990; Bajorath et al., 1991). Therefore, we investigated the electrostatic properties of the model as well as those of the energy-minimized kallikrein X-ray structure. Kallikrein at neutral pH carries a net charge of -13.5 and PSA, $+5.0$. The isopotential electrostatic surfaces (at -1 kcal/mol and $+1$ kcal/mol) of the minimized structures of kallikrein and PSA are shown in Figures 5B and 6B. Interestingly, despite the difference in net charge between kallikrein and PSA, the negative contour seems to be similar in the immediate vicinity of the catalytic triad residues. As expected for a protein with a net charge of -13.5 , kallikrein is largely covered by a negative potential (Fig. 5B). PSA was less covered by the negative potential and, relative to kallikrein, has larger positive potential surfaces surrounding the entrance of the active

Fig. 6. Electrostatic surface potential of PSA. **A:** PSA is represented in the same orientation as is kallikrein in Figure 5 with the same color code. **B:** In a similar orientation, isopotential contours for PSA are shown as for kallikrein in Figure 5B. PSA has a net charge of $+5$. The positive potential at $+1$ kcal/mol surrounds the entrance to the active site and part of the structure. Similarly to kallikrein, the negative contour at -1 kcal/mol encloses the catalytic triad residues. This result for the active-site triad was also found when analyzing the electrostatic surface potential of trypsin, chymotrypsin, and tonin (data not shown).

site and other parts of its structure (Fig. 6B). This result suggests that PSA is likely to interact with substrates very different from those of kallikrein.

Calculations of the electrostatic potential were also performed on the PSA model with the kallikrein loop conformation of lowest potential energy, which fully closes the active site. The results indicate that the active-site residues of PSA were still enclosed in a negative potential surface. Soman et al. (1989) have pointed out that negative potential in the vicinity of the active site is conserved between cow and rat trypsin despite a net charge difference of 12.5 units. This conservation might be necessary in the serine protease family in order to maintain the correct pK_a for the catalytic triad residues and therefore the enzymatic activity. We found similar results when comparing the electrostatic potential in the immediate region of the active site of the 4 reference X-ray proteins (data not shown). In all of these enzymes, the catalytic triad is enclosed in the negative part of the electrostatic isopotential surface (at -1 and -2 kcal/mol), although significant differences in the overall net charge of each protease are observed.

Asn 61 in the PSA structure, which is glycosylated in vivo (Schaller et al., 1987), is fully accessible to the solvent and located in a region of a positive potential. This result confirms our model structure because carbohydrates are negatively charged and their interaction with protein might be optimal in areas that are highly positively charged. The residues mainly responsible for this positive surface around Asn 61 are Arg 36, 39, 60, and 95g, and Lys 62 and 95e.

Active-site analysis

One important aim of our modeling study was to build the region of the substrate binding site of PSA. The active site and specificity pocket of PSA were analyzed based on the docking of bovine pancreatic trypsin inhibitor (BPTI) residues P3 to P'3 into the model structure. The loops around the active site form a circular wall in the PSA model similar to that found in the kallikrein structure. This "wall" is more pronounced than in trypsin or chymotrypsin and it might play a role in the substrate specificity. Hedstrom et al. (1992) have induced chymotrypsin-like activity for ester hydrolysis in trypsin by mutation of loop regions somewhat outside the active site that are not directly involved in the S1 binding pocket. This suggests that, although PSA might maintain the very conserved structure common to serine proteases, small shifts in the structure of some surface loops due to amino acid substitutions could lead to considerable changes in substrate specificity.

Analysis of the active-site region of the PSA model led to the following results. The residues involved in the catalytic triad in kallikrein (His 57, Asp 102, Ser 195) showed an RMS deviation between the initial and minimized structures of 0.25 Å for the backbone atoms and 0.34 Å for all atoms including hydrogens. The same calculations for PSA showed an RMS deviation of 0.30 Å and 0.56 Å, respectively. The bottom of the specificity pocket contained Ser 189 as does chymotrypsin; in contrast, this residue is an aspartate in tonin, kallikrein, and trypsin, and this negative charge determines these enzymes' P1 specificity for lysine or arginine residues. Thus, the model predicts that PSA would not interact very efficiently with a charged residue at the P1 position of a substrate. Gly 216 of PSA is conserved relative to chymotrypsin and trypsin and might allow for the pres-

ence of a bulky side chain at P1. However, PSA has a serine instead of a glycine at position 226, which could disrupt the interaction of PSA with some hydrophobic residues at P1, although this serine residue does not block off the S1 pocket.

To assess potential PSA-substrate interactions, the structure of the complex of kallikrein with BPTI (Chen & Bode, 1983) was used for further modeling studies. When introducing BPTI residues P3 to P'3 in the active site of PSA, using BPTI-kallikrein as a template, the characteristic hydrogen bonding pattern between the main-chain atoms of the substrate and of the serine protease enzyme was conserved (Fig. 7). These interactions included the backbone carbonyl of the P3 residue (Ile) with the peptide amide of Gly 216 of PSA. The hydrogen bond between the amide hydrogen of the P1 tyrosine residue with the backbone carbonyl of Ser 214 was slightly too long, but no attempt was made to further refine the structure at this level of investigation. The backbone carbonyl of residue P1 interacted with the amide hydrogens of Gly 193 and Ser 195. Finally, the amide hydrogen of the P'2 residue formed a hydrogen bond with the backbone carbonyl of Val 41. The side chains of BPTI residues P3 to P'3 were computationally mutated to the sequence IQYTYH, corresponding to the sequence of semenogelin I cleaved by PSA (Lilja et al., 1989). Examination of possible interactions of these side chains with PSA indicated that the hydrophobic P3 residue, isoleucine, was likely to interact with Trp 215 and His 172. Gln 174 was nearby, thus offering the possibility for a polar residue at P3 to interact favorably with PSA. A glutamine P2 residue might interact with the side chains of Ser 99 and Asp 97. Residue 99 is a tyrosine in kallikrein and interacts with the P2 residue, cysteine, of BPTI. Chymotrypsin has an isoleucine at position 99 and a preference for hydrophobic residues at the P2 position. The P1 residue, tyrosine, was accommodated in the S1 specificity pocket of PSA. The size of the cavity seems large enough for a large hydrophobic side chain possibly due to Pro 219. Pro 219 in kallikrein is conserved in PSA and located near the bottom of the specificity pocket. This proline residue assumes a *cis* conformation in kallikrein and contributes to enlargement of the binding pocket (Chen & Bode, 1983). This feature was conserved in the PSA model structure, and as a consequence, it is predicted that substrates having P1 residues with short side chains would not fill the S1 pocket and hence would be poor substrates. The P'1 threonine residue had van der Waals contacts with the disulfide bridge between Cys 42 and Cys 58. The P'2 residue tyrosine appeared to have aromatic-aromatic interactions with Phe 149 and hydrophobic contact with the carbon atoms of the aliphatic side chain region of Arg 39. The P'3 histidine residue interacted with the hydrophobic side chains of Val 41 and Leu 95c, suggesting the P'3 residue should be neutral and perhaps hydrophobic.

Structural changes are likely to occur during binding of a substrate to PSA, as shown for example in the case of the binding of BPTI to kallikrein (Chen & Bode, 1983), and thus our analyses of PSA-substrate interactions are preliminary.

Identification of a potential zinc ion binding site in the model structure of human PSA

Because human PSA is inhibited by Zn^{2+} ions (Watt et al., 1986), we therefore investigated the model structure in order to locate a zinc ion binding site. Among the structures of the 4 reference proteins, only tonin contains a zinc ion. The ligands di-

Fig. 7. Scheme for interactions of PSA with a putative substrate sequence of 6 amino acids. The P3 to P'3 IQYTYH residues of semenogelin I were docked in the active site of PSA as described in the Methods. The characteristic H bond pattern between the main-chain atoms of the substrate's sequence and the serine protease's active site were conserved in the model. The results suggest that several specific electrostatic and hydrophobic interactions may be possible or preferable. From the PSA model, it is predicted that the P3 residue is preferably hydrophobic or polar, that the P2 residue is preferably polar, and that the P1 residue can have a large and hydrophobic side chain inserted into the S1 pocket or could be polar. The P'1 residue could have a longer side chain and be hydrophobic, the P'2 residue can be hydrophobic or polar, and the P'3 residue should be hydrophobic.

rectly involved in Zn^{2+} ion binding in tonin are the NE2 atoms of His 57, 97, and 99 and the carboxyl group of Glu 148 from a neighboring molecule (Fujinaga & James, 1987). The binding of the Zn^{2+} ion appears to distort the active-site residue His 57 in tonin, thereby inhibiting the enzyme. The PSA model was inspected for potential histidine residues that could form a Zn^{2+} ion binding site. We suggest that His 91, 101, and 233 are likely to be involved in the binding of Zn^{2+} (Fig. 8). No significant movement of the α -carbon backbone of the PSA model was necessary to permit these 3 imidazole side chains to interact with the Zn^{2+} ion. The binding of a Zn^{2+} ion in this part of the molecule may modify the conformation of the loop 95 and could bring 1 of the aspartate residues (Asp 95, 97, or 98) in proximity to the 3 histidine residues, thus providing a fourth ligand to stabilize Zn^{2+} binding. Furthermore, Zn^{2+} ion binding could distort the catalytic triad through minor movements of His 101, which immediately precedes the active-site residue, Asp 102. The nearby tyrosine 234 may also contribute in the stabilization of the bound Zn^{2+} ion.

The sequence of rhesus macaque PSA (rmPSA) published by Gauthier et al. (1993) was compared to the human PSA sequence used for the model. These 2 sequences share about 89.5% identity; interestingly, the histidine 233 (in the chymotrypsin numbering system) of rmPSA is not conserved and is replaced by arginine, thus possibly altering the zinc ion binding site. However, overall the rmPSA should adopt a conformation very similar to that of human PSA.

Conclusion

A 3D model of the PSA enzyme has been constructed based on the crystallographic structures of the homologous serine proteases kallikrein and tonin because of the high sequence homology among PSA, kallikrein, and tonin. The resulting PSA model exhibits the typical topological motif and fold of the serine protease family and indicates conservation of characteristic inter-domain and intradomain interactions (e.g., ion pairs, disulfide bonds, hydrophobic clusters) as well as maintenance of the neg-

Fig. 8. Potential Zn^{2+} ion binding site in the human PSA model structure. Simplified representation of part of PSA showing the environment of a potential Zn^{2+} ion binding site. A Zn^{2+} ion was manually introduced in the model structure of PSA without energy refinements. In a ribbon representation, only the C-terminal helix and the loop 95 of PSA are shown. The side chains of His 91, 101, and 233 are likely to be involved in the binding of a Zn^{2+} ion. The side chain of the histidine residues putatively involved in Zn^{2+} ion binding are shown in solid black and the catalytic triad residue side chains (His 57, Asp 102, and Ser 195) appear stippled (gray). The location of the zinc ion, depicted as a sphere, is indicated by the arrow.

ative electrostatic potential surface near the active-site triad. The glycosylation site at Asn 61 is accessible to the solvent. The extended loop at residue 95 of PSA was modeled using molecular mechanics and dynamics. Solvation energy was taken into account to select a possible loop conformation; however, it also seems reasonable to suggest that this loop at residue 95 is quite flexible and may assume both "open" and "closed" conformations, which could be influenced by the binding of substrates or cofactors. The electrostatic surface potential contours of PSA, kallikrein, tonin, trypsin, and chymotrypsin were in agreement with a remarkable conservation of the negative electrostatic potential around the catalytic triad residues in the serine proteases.

PSA has been shown to discriminate among short oligopeptide substrates that are known to be cleaved by kallikrein or chymotrypsin (Christensson et al., 1990). This discrimination can be explained by the structures of surface loops and of the S1 pocket flanking the active site. The size of the PSA S1 specificity pocket is large enough to accommodate a large hydrophobic side chain. This is partly due to Pro 219, which enlarges the S1 cavity, and partly due to the presence of nearby amino acid residues with short side chains (Gly 216 and Ser 189). Residue 226, which is part of the S1 binding pocket, is a Ser in PSA and a Gly in chymotrypsin, and Ser 226 might also confer a particular specificity to PSA relative to chymotrypsin. A potential Zn^{2+} ion binding site in the human PSA model structure could involve the imidazole side chains of histidine residues 91, 101, and 233.

Methods

3D structural alignment

The coordinates of the X-ray structures of tonin (entry 1TON, Fujinaga & James, 1987), kallikrein (entry 2PKA, Bode et al.,

1983), chymotrypsin (entry 5CHA, Blevins & Tulinsky, 1985), and trypsin (entry 1TLD, Bartunik et al., 1989) were obtained from the Protein Data Bank (Bernstein et al., 1977). α -Carbon backbones of the 4 structures were superimposed interactively with the molecular graphics programs Insight II and Homology (Biosym Technologies, Inc.) starting at the catalytic triad residues while monitoring simultaneously the structural alignment (RMS) and the sequence homology. All deviations less than 0.8 Å were designated SCRs, independently of whether the region belonged to a secondary structure element or to a loop region. Only SCRs common to all 4 reference proteins were used for the construction of the initial model (Fig. 1). The normalized solvent accessibilities (Lee & Richards, 1971) of the side chains were investigated for the 4 reference proteins to identify the buried regions that form the hydrophobic cores (Fig. 1). All side chains with a surface accessibility less than 0.20 Å² were considered to be internal.

Sequence alignment

The Needleman and Wunsch scoring algorithm (Needleman & Wunsch, 1970) with the Dayhoff matrix (Dayhoff et al., 1983) (gap penalty and matrix bias of 3) was used as a starting point to compare the amino acid sequences of PSA (Lundwall & Lilja, 1987; Schaller et al., 1987) and kallikrein. Then the PSA sequence was aligned by eye using information from the structural alignment. This procedure avoided, when possible, introduction of gaps or insertions in highly conserved structural regions (Greer, 1990). PSA, like tonin and kallikrein, has an insertion at residue 95 compared to trypsin and chymotrypsin. This insertion in PSA, referred to as the "kallikrein loop," involves 8 residues (LLKNRFLR). Tonin and kallikrein also have an insertion in this part of the molecule, but no structural information is available for this region in either structure (Bode et al., 1983; Fujinaga & James, 1987). A deletion of 2 residues in the PSA sequence compared to kallikrein occurs in the loop region 145 (Fig. 1). Tonin has a loop of the same length and was used for the modeling of PSA in this area.

Construction of the model

The model was constructed from the atomic coordinates of tonin and kallikrein with the programs InsightII and Homology (Biosym Technologies, Inc.). The atomic coordinates of kallikrein were used for most of the model because its sequence homology was highest (Schaller et al., 1987). The general model building was done as follows: where possible, values for the main-chain and side-chain torsion angles (χ) of kallikrein were used for each substitution. If the PSA side chain was longer than the kallikrein side chain, an extended conformation was used for the remaining χ angles. Two regions of the PSA model were built with the crystallographic coordinates of the tonin enzyme as follows. For the kallikrein loop of PSA, coordinates from tonin residues Ile 95–Leu 95b (IPL) and Pro 95k–His 99 (PVHDH) were used to build the regions preceding and following this loop in the model. In the tonin structure, a gap of about 8 Å is observed because the electron density was not interpretable for this particular segment. The equivalent intervening loop segment (95cLLKNRFLR95j) of PSA was constructed within the Homology program by using randomly generated values for ψ and ϕ angles that fit the distance between the 2 anchor residues (Ser 95c

and Pro 95k) in the PSA model (Shenkin et al., 1987). The newly built loop was energy minimized by 200 cycles of steepest descents algorithm, while the rest of the PSA structure was held fixed. For the PSA sequence Pro 146 to Leu 150 (PEEFL), tonin was used as a template. Kallikrein has 8 residues in this loop region, whereas tonin and PSA have 6. This region was, as in the case of the kallikrein loop, minimized independently using several cycles of steepest descents minimization.

Coordinates for the modeled structure have been deposited in the Brookhaven Protein Data Bank (1PFA).

Energy minimization

No severe atomic collisions were detected during the construction of the PSA model. The minor steric clashes and bond strains due to the loop building and the introduction of the new side chains were removed using the molecular simulation program Discover (Biosym Technologies, Inc.) and the corresponding CVFF force field parameters (Hagler et al., 1974; Hagler & Lifson, 1974) on a Silicon Graphics workstation R4000-50/GTX. Hydrogen atoms were added to the model structure according to standard geometries. The refinement of the structure was carried out in stages (Mackay et al., 1989). The first step was performed using a high tethering force constant ($K = 1,000$ kcal/mol/Å²) on the protein heavy atoms and minimized with the steepest descents algorithm until the maximum Cartesian derivative was less than 10 kcal/mol/Å. This enabled relaxation of all newly positioned atoms without significant disruption of the overall structure. Then 3 additional rounds of minimization were done with restraints on the protein heavy atoms as follows: first, $K = 100$ kcal/mol/Å² until the maximum derivative was less than 5 kcal/mol/Å using the steepest descents algorithm; second, $K = 50$ kcal/mol/Å² until the maximum derivative was less than 3 kcal/mol/Å; and third, $K = 25$ kcal/mol/Å² until the maximum derivative was less than 2 kcal/mol/Å using conjugate gradients. Finally, unconstrained energy minimization was carried out until the maximum derivative was less than 1 kcal/mol/Å. All calculations were performed using a 20-Å cutoff for nonbonded interactions and a dielectric constant of 1. Because solvent molecules were not included, all the amino acids were defined to be in their neutral states (Mackay et al., 1989) to avoid artifactual dominance of electrostatic terms. In order to provide a control by which to judge the results of the calculations for the PSA model, energy minimization of the kallikrein X-ray structure was carried out in parallel using the same strategy.

Refinement of the kallikrein loop of PSA

A conformational search using a 50-ps high-temperature dynamics simulation was performed on the kallikrein loop of the PSA minimized model. The 16 residues between the SCRs flanking this loop region were allowed to move while the rest of the structure was held fixed. The leapfrog integration algorithm (Hockney & Eastwood, 1981) with a 1-fs time step was used. Initial velocities for the simulation were assigned from a Maxwellian distribution at a temperature of 1,000 K. In order to avoid possible *trans-cis* transitions due to the high-temperature simulation, a torsional restraint of 5 kcal mol⁻¹ rad⁻² was added to all peptide bonds, except for the proline residue, for which the restraint forced the bond to be either *cis* or *trans* (the angle was

forced to whichever it was closer to). The nonbond cutoff was 20 Å and a history file was written every picosecond. Each structure resulting from the simulation was minimized in parallel using 100 cycles of steepest descents algorithm with only the region between the 2 SCRs flanking the loop allowed to move. The 50 minimized conformations were then annealed to 300 K for 5 ps. Finally, these structures were further minimized with 100 steps of steepest descents and 2,000 steps of conjugate gradients. An energetic criterion was used to analyze the result further. All the above calculations were carried out in vacuo, and the amino acids were defined to be in their neutral states to keep charged side chains from dominating the structural search.

The solvation energy calculation is described below.

Electrostatic calculations

The finite difference Poisson-Boltzmann (FDPB) method (Klapper et al., 1986; Gilson et al., 1988) was used for the evaluation of the solvation energies and for the calculation of the 3D distribution of the electrostatic potential. The FDPB was solved with the program DelPhi (Biosym Technologies, Inc.). The electrostatic potential can be described by the linearized Poisson-Boltzmann equation (McQuarrie, 1976):

$$\nabla \epsilon(r) \nabla \phi(r) - \epsilon(r) k^2(r) \phi(r) = -4\pi \rho(r),$$

where ϕ is the electric potential due to the charge density ρ , ϵ is the dielectric constant, and k is the inverse salt screening length. The dielectric constant takes the value ϵ_m in the protein interior and ϵ_s in the solvent. The solvation energy can be evaluated by 2 calculations with identical grid mappings and the same interior dielectric. In the first calculation, the exterior dielectric constant ϵ_s was set to 80 and ϵ_m set to 4; for the second, ϵ_s was set to 1 and ϵ_m kept at 4. The difference in the total energy for these 2 calculations corresponds to the change in electrostatic energy in going from vacuum to solvent, which is the solvation energy (Born, 1920; Gilson & Honig, 1988):

$$\Delta G_{\text{solv}} = 1/2 \sum_i q_i (\phi_i^{\text{solvent}} - \phi_i^{\text{vacuum}})$$

where $\phi_i^{\text{solvent}} - \phi_i^{\text{vacuum}}$ is the change in calculated potential at charge q_i upon altering the solvent. For this calculation, atoms were assigned the partial charges from the CVFF force field, with the side chains of Arg, Lys, Glu, and Asp assumed to be charged.

Isopotential surfaces were calculated to compare the electrostatic properties of the minimized X-ray structure of kallikrein with that of the PSA model. The ionic strength was set to 0.1 and the dielectric constant to 4 (Harvey, 1988) for the protein interior and 80 for the surrounding solvent. The atomic coordinates were mapped onto a cubic 3D grid of $65 \times 65 \times 65$ Å, with a distance of 10 Å between the protein and the grid edge, and a 1-Å/grid point. Formal charges were assigned to the ionized groups (Soman et al., 1989): Arg, Lys, and the N-terminal α -amino group were each assigned a charge of +1; Asp, Glu, and the C-terminal carboxyl group a charge of -1; and His a charge of 0.5 distributed equally between the 2 imidazole ring nitrogen atoms except for the active-site His 57, which was considered to be neutral (enzyme in the resting state; Kos-

siakoff & Spencer, 1981). The resulting 3D electrostatic maps were analyzed with the graphics program Insight II.

Ligand docking

To explore potential binding interactions between the PSA structure and a small peptide substrate, we have positioned the BPTI in the active site of the model. The BPTI complexed with kallikrein (Chen & Bode, 1983) was superimposed onto the PSA model and was used as a first template. The inhibitor reactive site and the enzyme active site are referred to as $Pn \dots P1 \dots P'1 \dots P'n$ and $Sn \dots S1 \dots S'n$, respectively, according to Schechter and Berger (1967). The reactive bond is located between $P1$ and $P'1$. As pointed out by Greer (1981) (Bode & Huber, 1992), the serine protease inhibitors have a highly conserved structure for the backbone atoms between residues $P2$ to $P'2$ and start to differ slightly after residue $P3$ and $P'3$. Therefore, only residues $P3$ to $P'3$ of BPTI were selected and computationally mutated according to one peptide sequence (IQYTYH) of semenogelin I suggested to be cleaved by PSA (Lilja et al., 1989). Then a short energy minimization (200 cycles of steepest descents with a tethered force constant of 100 kcal/mol/Å²) was carried out on the PSA model complexed with the IQYTYH peptide structure in order to refine partially the active-site area. Structural analysis of the active-site region and evaluation of possible hydrogen bond and hydrophobic interactions between PSA and this peptide were investigated.

Acknowledgments

We thank Dr. Moises Hassan, from Biosym Technologies, Inc., for helpful suggestions. We also thank Drs. Judith Greengard and Cindy Fisher for critical reading of the manuscript, and Drs. Mary J. Heeb, Ben Gittes, and John Tainer for helpful discussions.

References

- Akiyama K, Nakamura T, Iwanaga S, Hara M. 1987. The chymotrypsin-like activity of human prostate-specific antigen, gamma-seminoprotein. *FEBS Lett* 225:168-172.
- Bajorath J, Kitson DH, Kraut J, Hagler AT. 1991. The electrostatic potential of *Escherichia coli* dihydrofolate reductase. *Proteins Struct Funct Genet* 11:1-12.
- Bartunik HD, Summers LJ, Bartsch HH. 1989. Crystal structure of bovine β -trypsin at 1.5 Å resolution in a crystal form with low molecular packing density. Active site geometry, ion pairs and solvent structure. *J Mol Biol* 210:813-828.
- Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasmui M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol* 112:535-542.
- Blevins RA, Tulinsky A. 1985. The refinement and the structure of the dimer of α -chymotrypsin at 1.67 Å resolution. *J Biol Chem* 260:4264-4275.
- Blundell TL, Sibanda BL, Sternberg MJE, Thornton JM. 1987. Knowledge-based prediction of protein structures and the design of novel molecules. *Nature* 326:347-352.
- Bode W, Chen Z, Bartels K, Kutzbach C, Schmidt-Kastner G, Bartunik H. 1983. Refined 2 Å X-ray crystal structure of porcine pancreatic kallikrein A, a specific trypsin-like serine proteinase. *J Mol Biol* 164:237-282.
- Bode W, Huber R. 1991. Proteinase-protein inhibitor interaction. *Biomed Biochim Acta* 50:437-446.
- Bode W, Huber R. 1992. Natural protein proteinase inhibitors and their interaction with proteinases. *Eur J Biochem* 204:433-451.
- Born M. 1920. Volumes and heats of hydration of ions. *Z Physik* 1:45-48.
- Bowie JU, Luthy R, Eisenberg D. 1991. A method to identify protein sequences that fold into a known three-dimensional structure. *Science* 255:164-170.
- Brawer MK. 1991. Prostate specific antigen: A review. *Acta Oncol* 30:161-168.
- Chen Z, Bode W. 1983. Refined 2.5 Å X-ray crystal structure of the complex formed by porcine kallikrein A and the bovine pancreatic trypsin inhibitor. *J Mol Biol* 164:283-311.
- Christensson A, Laurell CB, Lilja H. 1990. Enzymatic activity of prostate-specific antigen and its reactions with extracellular serine proteinase inhibitors. *Eur J Biochem* 194:755-763.
- Cohen P, Graves HCB, Peehl DM, Kamarei M, Giudice LC, Rosenfeld RG. 1992. Prostate-specific antigen (PSA) is an insulin-like growth factor binding protein-3 protease found in seminal plasma. *J Clin Endocrinol Metab* 75:1046-1053.
- Dayhoff MO, Barker WC, Hunt LT. 1983. Establishing homologies in protein sequences. *Methods Enzymol* 91:524-545.
- Fine RM, Wang H, Shenkin PS, Yarmush DL, Levinthal C. 1986. Predicting antibody hypervariable loop conformations II: Minimization and molecular dynamics studies of MCPC603 from many randomly generated loop conformations. *Proteins Struct Funct Genet* 1:342-362.
- Fisher CL, Greengard JS, Griffin JH. 1994. Models of the serine protease domain of the human antithrombotic plasma factor activated protein C and its zymogen. *Protein Sci* 3:588-599.
- Fujinaga M, James MNG. 1987. Rat submaxillary gland serine protease, tonin, structure solution and refinement at 1.8 Å resolution. *J Mol Biol* 195:373-396.
- Furie B, Bing DH, Feldmann RJ, Robinson DJ, Burnier JP, Furie BC. 1981. Computer-generated models of blood coagulation factor Xa, factor IXa, and thrombin based upon structural homology with other serine proteases. *J Biol Chem* 257:3875-3882.
- Gauthier ER, Chapdelaine P, Tremblay RR, Dubé JY. 1993. Characterization of rhesus monkey prostate specific antigen cDNA. *Biochim Biophys Acta* 1174:207-210.
- Gilson M, Honig B. 1988. Calculation of the total electrostatics energy of a macromolecular system: Solvation energies, binding energies and conformational analysis. *Proteins Struct Funct Genet* 4:7-18.
- Gilson M, Sharp K, Honig B. 1988. Calculating the electrostatic potential for molecules in solution. *J Comp Chem* 9:327-335.
- Gittes RF. 1991. Carcinoma of the prostate. *N Engl J Med* 324:236-243.
- Grave HCB, Sensabaugh GF, Blake RT. 1985. Postcoital detection of male-specific semen protein. Application to the investigation of rape. *N Engl J Med* 312:338.
- Greer J. 1981. Model of a specific interaction. *J Mol Biol* 153:1043-1053.
- Greer J. 1985. Model structure for the inflammatory protein C5a. *Science* 228:1055-1060.
- Greer J. 1990. Comparative modeling methods: Application to the family of the mammalian serine proteases. *Proteins Struct Funct Genet* 7:317-334.
- Hagler AT, Huler E, Lifson S. 1974. Energy functions for peptides and proteins. I. Derivation of a consistent force field including the hydrogen bond from amide crystals. *J Am Chem Soc* 96:5319-5327.
- Hagler AT, Lifson S. 1974. Energy functions for peptides and proteins. II. The amide hydrogen bond and calculation of amide crystal properties. *J Am Chem Soc* 96:5327-5335.
- Hara M, Inorre T, Fukuyama T. 1971. Some physico-chemical characteristics of gamma-seminoprotein, an antigenic component specific for human seminal plasma. *Jpn J Legal Med* 25:322.
- Harvey S. 1989. Treatment of electrostatic effects in macromolecular modeling. *Proteins Struct Funct Genet* 5:78-92.
- Hedstrom L, Szilagyi L, Rutter WJ. 1992. Converting trypsin to chymotrypsin: The role of surface loops. *Science* 255:1249-1253.
- Hockney RW, Eastwood JW. 1981. *Computer simulation using particles*. New York: McGraw-Hill.
- Jones TA, Thirup S. 1986. Using known substructures in protein model building and crystallography. *EMBO J* 5:819-822.
- Klapper I, Hagstrom R, Fine R, Sharp K, Honig B. 1986. Focusing of electric fields in the active site of Cu-Zn superoxide dismutase: Effects of ionic strength and amino-acid modification. *Proteins Struct Funct Genet* 1:47-59.
- Kossiakoff AA, Spencer SA. 1981. Direct determination of the protonation states of aspartic acid-102 and histidine-57 in the tetrahedral intermediate of the serine proteases: Neutron structure of trypsin. *Biochemistry* 20:6462-6474.
- Lee B, Richards FM. 1971. The interpretation of protein structure: Estimation of static accessibility. *J Mol Biol* 55:379-400.
- Li TS, Beling CG. 1973. Isolation and characterization of two specific antigens of human seminal plasma. *Fertil Steril* 24:134.
- Lilja H. 1985. A kallikrein-like serine protease in prostatic fluid cleaves the predominant seminal vesicle protein. *J Clin Invest* 76:1899-1903.

- Lilja H, Abrahamsson PA, Lundwall A. 1989. Semenogelin, the predominant protein in human semen. Primary structure and identification of closely related proteins in the male accessory sex glands and on the spermatozoa. *J Biol Chem* 264:1894-1900.
- Lilja H, Cockett ATK, Abrahamsson PA. 1992. Prostate specific antigen predominantly forms a complex with α 1-antichymotrypsin in blood. *Cancer* 70:230-234.
- Lundwall A, Lilja H. 1987. Molecular cloning of human prostate specific antigen cDNA. *FEBS Lett* 214:317-322.
- Mackay DHJ, Cross AJ, Hagler AT. 1989. *The role of energy minimization in simulation strategies of bimolecular systems*. New York: Plenum.
- McQuarrie DA. 1976. *Statistical mechanics*. New York: Harper and Row.
- Needleman SB, Wunsch CD. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* 48:443-453.
- Novotny J, Rashin AA, Brucoleri RE. 1988. Criteria that discriminate between native proteins and incorrectly folded models. *Proteins Struct Funct Genet* 4:19-30.
- Oesterling JE. 1991. Prostate specific antigen: A critical assessment of the most useful tumor marker for adenocarcinoma of the prostate. *J Urol* 145:907-923.
- Rashin AA, Honig B. 1984. On the environment of ionizable groups in globular proteins. *J Mol Biol* 173:515-521.
- Riegman PHJ, Klassen P, van der Korput JAGM, Romijn JC, Trapman J. 1988. Molecular cloning and characterization of novel prostate antigen cDNA's. *Biochem Biophys Res Commun* 155:181-188.
- Schaller C, Akiyama K, Tsuda R, Hara M, Marti T, Rickli EE. 1987. Isolation, characterization and amino-acid sequence of γ -seminoprotein, a glycoprotein from human seminal plasma. *Eur J Biochem* 170:111-120.
- Schechter I, Berger A. 1967. On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* 27:157-162.
- Sensabaugh GF. 1978. Isolation and characterization of a semen-specific protein from human seminal plasma: A potential new marker for semen identification. *J Forensic Sci* 23:106.
- Sharp K, Honig B. 1990. Electrostatic interactions in macromolecules: Theory and applications. *Annu Rev Biophys Chem* 19:301-332.
- Shenkin PS, Yarmush DL, Fine RM, Wang H, Levinthal C. 1987. Predicting antibody hypervariable loop conformation I. Ensembles of random conformations for ringlike structures. *Biopolymers* 26:2053-2085.
- Soman K, Yang AS, Honig B, Fletterick R. 1989. Electrical potentials in trypsin isozymes. *Biochemistry* 28:9918-9926.
- Stamey TA. 1989. Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate. *Monograph Urol* 10:49.
- Struthers RS, Kitson DH, Hagler AT. 1991. Predicted three-dimensional structure of the protease inhibitor domain of the Alzheimer's disease β -amyloid precursor. *Proteins Struct Funct Genet* 9:1-11.
- Summers NL, Karplus M. 1990. Modeling of globular proteins, a distance-based data search procedure for the construction of insertion/deletion regions and Pro-non-Pro mutations. *J Mol Biol* 216:991-1016.
- Wang MC, Valenzuela LA, Murphy GP, Chu TM. 1979. Purification of a human prostate specific antigen. *Invest Urol* 17:159.
- Watt KWK, Lee PJ, Timkulu TM, Chan WP, Loo R. 1986. Human prostate-specific antigen: Structural and functional similarity with serine protease. *Proc Natl Acad Sci USA* 83:3166-3170.