ELSEVIER

Construction of a 3D model of oligopeptidase B, a potential processing enzyme in prokaryotes

Tímea Gérczei,† György M. Keserü,*,† and Gábor Náray-Szabó†

**Department of Chemical Information Technology, Technical University of Budapest, Budapest, Hungary*
†Department of Theoretical Chemistry, Loránd Eötvös University, Budapest, Hungary

A three dimensional structural model of oligopeptidase B (OpB) was constructed by homology modeling. High resolution X-ray structure of prolyl oligopeptidase (PEP), the only protein with sequential and functional homology was used as a template. Initial models of OpB were built by the MODELLER and were analysed by the PROCHECK programs. The best quality model was chosen for further refinement by two different techniques—either constrained molecular dynamics simulations or simulated annealing calculations starting from 500 K. The overall quality of each of the refined models was evaluated and the simulated annealing procedure found to be more effective. The refined model was analysed by different protein analysis programs including PROCHECK for the evaluation of the Ramachandran plot quality, PROSA for testing interaction energies and WHATIF for the calculation of packing quality. This structure was found to be satisfactory and also stable at room temperature as demonstrated by a 300 ps long unconstrained molecular dynamics simulation. Calculation of molecular electrostatic potentials revealed that the binding site of OpB is more negative than that of PEP, in accordance with the experimentally observed selectivity of OpB towards proteolysis at dibasic sites. A recently developed Monte Carlo docking method was used provide a structural rationale for the affinity differences measured between Z-Arg and Z-Arg-Arg substrates. © *2000 by Elsevier Science Inc.*

INTRODUCTION

The long search for prohormone convertases and neuropeptide precursors has already been successful in eukaryotes. These processing enzymes cleave at the carboxy end of dibasic sites (Arg-Arg or Lys-Arg) and have mostly an amino acid composition similar to the serine protease subtilisin.[1] Since OpB—the target of the present study—cleaves peptides at dibasic sequence two order of magnitudes faster[2,3] than at monobasic ones, it could be a good candidate for a similar convertase in prokaryotes, and suggests that these proteases are of ancient origin.

OpB—previously called Protease II—belongs to the new prolyl oligopeptidase family[4,5] of serine proteases and cleaves peptides consisting of no more than 30 amino acids. At present the only known structure from this family is the recently obtained X-ray structure of Pep (PDB ID code 1QFM and 1QFS).[6,7] This structure showed that these proteins have two domains: a so-called β-propeller domain which excludes proteins from the enzyme and saves them from proteolytic damage and a catalytic domain with an α/β hydrolase fold typical for serine proteases. The catalytic activity of OpB is strongly dependent on ionic strength not typical for serine proteases, reflecting a strong electrostatic connection between the dibasic substrate and a negatively charged group (carboxyl group) of the enzyme.[3]

The homology model for OpB was generated using the 3D structure of Pep based on a sequence identity of 29% and a sequence homology of 42% between the two proteases. In order to understand the enhanced catalytic efficiency of OpB for dibasic sites versus monobasic ones two substrates—Z-Arg and Z-Arg-Arg (Z = benzyloxycarbonyl)—were docked into the active site of the model structure with a newly developed Monte Carlo docking method. The role of electrostatics in the binding of substrates was characterised by calculating the electrostatic potential of the binding site for the two enzyme-substrate complexes. Based on our findings, we attempt to explain the enhanced catalytic efficiency for dibasic substrates of OpB relative to PEP and the reduced activity in the case of a higher ionic strength.

The determination of the structure of OpB by X-ray crystallography is underway. An experimental structure will give

The Color Plates for this article are on pages 57–58.

Corresponding author: G. M. Keserü, Computer Assisted Drug Discovery, Gedeon Richter Ltd., H-1475 Budapest, P.O. Box 27, Hungary

E-mail address: gy.keseru@richter.hu

Journal of Molecular Graphics and Modelling 18, 7–17, 2000

655 Avenue of the Americas, New York, NY 10010

1093-3263/00/$–see front matter
PII S1093-3263(00)00042-X

the opportunity on the one hand for a quality check of our modelling method and on the other hand to produce more reliable models of the enzymes belonging to the Pep family.

METHODS

Sequence Alignment

Sequences of PEP and OpB isolated from pig (PPCE_PIG) and *E. coli* (PTRB_ECOLI), respectively were obtained from the Swiss-prot database[8] (entry codes P23687 and P24555 for PEP and OpB, respectively). Sequence alignment was derived using the CLUSTALW package,[9] default parameters were applied. Aligned sequences were then inspected and adjusted manually to minimize the number of gaps and insertions.

Rough Model

Sequence alignment between OpB and PEP was used to derive several 3D models for OpB. The first model has been constructed from the sequence alignment between OpB and PEP and the 3D structure of PEP[5] using the academic version 4.0 of MODELLER[10,11] with default parameters that proposed loop conformations. Molecular PDF values obtained after low level simulated annealing are collected to Table 1. The quality of proposed models (model A, B and C) was checked by PROCHECK[12] and secondary structures were assigned. Secondary structure of OpB was also predicted by the PhD program.[13] PROCHECK results and the extent of secondary structural elements in comparison with the PhD prediction were carefully analyzed for model A, B and C, respectively. Model A having the most extent secondary structural motifs was chosen for further refinement. Although the Ramachandran plot and the overall quality of Model B and C were better than those of Model A, the more extent secondary structural elements, found in Model A, justified its selection for refinement. The reason for this selection is that the quality of the model can be improved but new secondary structural elements usually cannot be generated during refinement.

Refinement

Model A, obtained by MODELLER, was extended by water molecules copied from the X-ray structure of PEP. The solvated protein was subjected to constraint energy minimization with a harmonic constraint of 100 kJ/mol/Å^2, applied for all protein atoms, using the steepest descent technique to eliminate bad contacts between protein atoms and structural water molecules. The resulted structure was finally refined by two different techniques. All calculations were performed by the CFF force field implemented to the X-PLOR package.[14] Energy minimizations were carried out by performing Powell conjugate gradient optimization until the maximum derivative became less than 0.01 kJ/mol/Å. Electrostatic treatment involved a distance dependent dielectric constant ($\varepsilon = 4R$) and atomic charges for all protein atoms obtained from the original CHARMM force field.[15]

We used a multi step refinement process involving a series of constrained molecular dynamics (MD) simulations at first. In this case, Model A was subjected to subsequent constrained simulations using constraints of 100, 35, 20 and 5 kJ/mol/Å^2 on all backbone atoms, respectively. The resulting structure was finally minimized without any constraint to yield Model A1 as a refined structure. Next we applied the popular simulated annealing procedure on the whole system using the simulation protocol described by Szilágyi and Závodszky.[16] Constrained simulation with 20 kJ/mol A^2 harmonic constraint on all backbone atom was started at 500 K using the SHAKE algorithm.[17] The temperature was decreased systematically in every step by 5%, until 300 K was reached. The system was simulated for 1 ps (420 fs using direct velocity scaling and 580 fs with coupling to the heat bath) at each temperature. A final unconstrained energy minimization to reach convergence with a gradient of 0.01 kJ/mol/Å was used to obtain Model A2. Testing the effect of starting temperature on the refinement, this procedure was repeated starting simulated annealing at 900 K and following the procedure described previously. This refinement led to Model A3. The quality of these refined models were examined using the PROCHECK program and Model A2 having the best overall quality was chosen for further evaluation.

Evaluation of Refined Models

In the last step of homology modeling the refined structure of Model A2 was subjected to a series of four tests for its internal consistency and reliability. Backbone conformation was evaluated by the inspection of the Psi/Phi Ramachandran plot obtained from PROCHECK analysis. The PROSA[18] test was applied to check for an energy criteria in comparison with the potential of mean force derived from a large set of known protein structures. Packing quality of the refined structure was investigated by the calculation of WHATIF Quality Control value.[19] Finally the dynamic behavior of Model A2 was investigated during 300 ps of unconstrained MD simulation to test its stability.

Electrostatic Calculations

As we wanted to give an explanation to the different binding affinity of the protein for its substrates, the electrostatic potential of the binding site cavity were calculated for both enzyme-substrate complexes with the DelPhi package[20,21] using the continuum dielectric approach to solve the Poisson-Boltzmann equation by the finite difference method. We used charges obtained from the AMBER force field for the protein, the substrates were not charged. Since the enzyme activity maximum is near pH = 8.00 at 25°C, we used charges (protonation states) belonging to this pH. Residues for which net charges differed from zero are: Asp(−1), Lys(+1), Arg(+1), Glu(−1), His(0.50). We used a constant ionic strength of 0.145 M in all calculations. The dielectric constants were, as usual, 4 for the protein and 80 for the solvent (water) environment, respectively. The structural water molecules were treated as part of the protein with a dielectric constant of 4.

Docking

The refined structure of Model A2 was used for docking calculations. All calculations were performed using the AMBER* force field[22] available in the MacroModel package.[23] This force field is identical to the original AMBER one with the exception of backbone parameters. Electrostatic treatment of the system involves a distance-dependent dielectric constant

(ε = 4R)[24] and AMBER all-atom charges for amino acid residues at this stage of calculations. A binding site model was defined as a sphere with an approximately 15 Å radius around the catalytically active Ser-554 residue identified in this class of serine proteases. This model is composed of amino acids which were not contiguous in the protein, therefore all the backbone atoms had to be constrained applying a 20 kJ/mol/Å^2 tethering function in the constrained system. Side-chain atoms and the substrate were allowed to move freely in the substructure system. Cut-off distances were extended to 8 Å for van der Waals and 15 Å for charge-charge interactions, respectively, considering long-range electrostatic interactions in the protein. The combined conformational search used in these docking calculations involves the variation of the internal degrees of freedom of the substrate and the binding site (rotatable bonds) as well as the external degrees of freedom of the ligand (relative translations and rotations) relative to the binding site. All of the internal and external degrees of freedom of the ligand and a limited set of internal degrees of freedom of the protein were considered.

Monte Carlo (MC) searches involve the random variation of all rotatable bonds combined with the so-called variable molecules selection (MOLS) for translations and rotations of the ligand in the binding site.[25,26,27] The MOLS routine allows a selected molecule to move during an MC step by randomly selected rotation and/or translation with respect to a fixed active site and, therefore, is particularly useful for docking calculations. The combined MCMM/MOLS procedure was applied to allow the simultaneous random translation and rotation of ligands with respect to the binding site with increments of 0.5–1.0 Å for the translation of the center of mass of the ligand along the *x*, *y*, and *z* axes and 30° to 180° for the rotation of the whole ligand around the same axes, respectively. This approach allowed us to look for all possible low energy binding modes of ligands in a single MC simulation.

The initial random structures of the complex generated by the combined MCMM/MOLS procedure were subjected to restrained energy minimization applying a quadratic restoring potential to the backbone atoms of the active site. The constraint-substructure shell system described previously was used during minimization and the resulting minimum energy complex structures were sorted by energy. The unique structures within a 50 kJ/mol energy window above the global minimum were stored in a MacroModel multi structure file for further interactive study using the three-dimensional graphical interface of MacroModel. Conformational analysis of the enzyme-substrate complex involved 1,000 MC steps (MCMM/MOLS) and unique binding conformations were stored after rejections based on the comparison of binding geometry of the substrate obtained in each step. Non-bonded atom pairs of extreme proximity give rise to very high van der Waals repulsion, therefore trial structures having such close pairs were also discarded before minimization. Low energy conformers obtained were analyzed comparing the orientation of the substrate relative to the catalytically active Ser-554.

RESULTS AND DISCUSSION

Homology Modeling

The final alignment of the OpB sequence to that of PEP is shown in Figure 1. Besides the insertions and deletions detected in loop regions corresponding to secondary structural units of PEP there are only a few gaps. The optimal alignment between OpB and PEP constrains four insertions in b6/4 (1 residue), b2 (1 residue), b8 (1 residue) and aD1 (3 residues) and four deletions in b1 (1 residues), b4/4 (2 residues), aA (1 residues) and aE (3 residues). The consequence of gaps in aD1, b8, b4/4, aA and aE is only the shortening of strands and helices since these insertions and deletions can be considered as non-breaking gaps. It is interesting that PhD also predicts shorter aD1 and aA helices in OpB than were detected in PEP. As our aim is to investigate structure-function relationships in OpB, the structure of its active site and the substrate binding pocket is of the utmost importance. The catalytic domain of PEP is built up of residues 1–72 and 428–710. The substrate binding site is arranged on the b5-b8 strands and aC and aE helices including the characteristic Ser-His-Asp catalytic triad located at positions 554, 680 and 641, respectively.

Inspection of this primary sequence alignment between Pep and OpB shows that the sequence identity in the catalytic domain is higher (30.17%) than that in the β-propeller domain (21.75%) and increases to 36% in the vicinity of the catalytic triad. Therefore, the most important part of the sequence for catalytic activity is mostly conserved. The amino acid composition of the two enzymes is virtually the same: the percentages of the apolar, polar and charged amino acids are 47.1%, 25.1% and 27.8% for OpB and 49.5%, 24.1% and 26.4% for Pep, respectively. The isoelectric points of the two enzymes are also similar (pI_{OpB} = 5.2[28] and pI_{Pep} = 4.9[29]). Considering the alignment between Pep and OpB, we can observe that the hydrophilic sequence segments which are near to the solvent and the hydrophobic segments which are buried in the 3D structure of Pep are conserved globally in OpB (Fig. 1). Although we could not identify any extremely charged sequence segments in the two enzymes, there is a hydrophobic segment near to the catalytic Ser (Ser-532) from Leu-525 to Ile-542 in OpB which has no equivalent in Pep.[30] In that segment of OpB there are some polar to apolar mutations versus Pep: Leu-525 mutated from Arg, Ala-533 from Asn, Met 529 from Asn and Val 540 from Thr. The reason for this may be the specificity of OpB toward basic substrates, so a positively charged amino acid like an Arg is unfavorable near to the catalytic site. It should be noted, however that the overall character of the aligned Pep sequence is also hydrophobic.

Our alignment was carefully checked in these regions and we found that all of the critical structural elements involved in the binding and catalytic transformation of the substrate are intact. Therefore we conclude that this alignment can be used to construct a reliable 3D model for OpB binding site.

The results of PROCHECK analysis obtained for the three best quality structures are collected in Table 1. Given the relatively low percentage of residues having disallowed torsional angles, the quality of Ramachandran plots is acceptable for all models. However, goodness factors calculated both for covalent and non-covalent interactions suggest that these models have to be refined. Molecular PDF values obtained by MODELLER also suggested that further refinement was required. Considering the computational expense of such a procedure a single structure was selected for further refinement that may improve the Ramachandran plot quality and goodness factors, however, leaving the folding pattern essentially unchanged. Thus, the extent of secondary structural elements was used as the basic selection criteria. Secondary structures of Model A, B and C, compared to predictions based on the

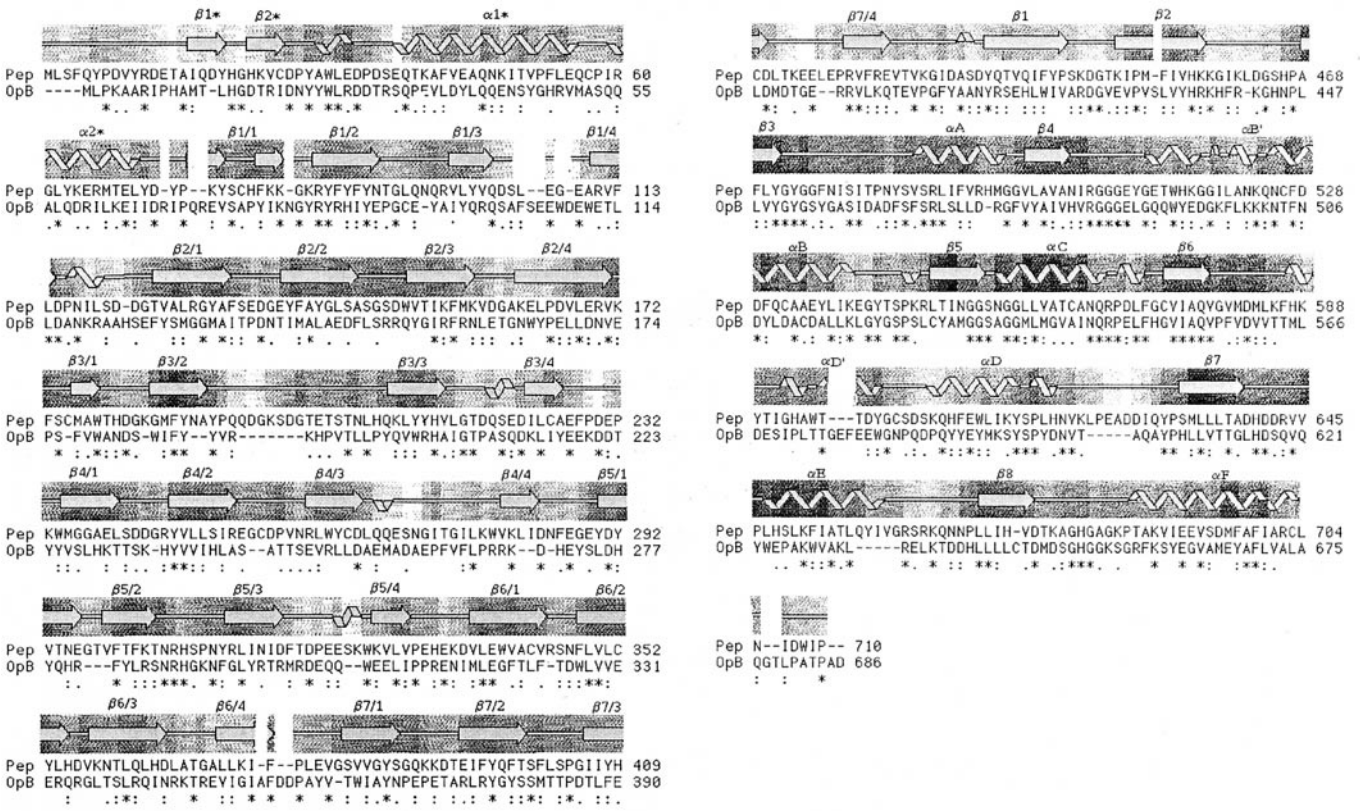


Figure 1. The sequence alignment between Pep and OpB produced by Clustal W. The secondary structure of Pep is shown with arrows showing β-sheets and spirals for α-helices. The background of the sequence alignment reflects the solvent accessibility of the amino acids in the Pep structure on a scale of black for buried to white for solvent accessible.

Table 1. Quality of Structures Predicted by the MODELLER Program[a]

	Ramachandran plot quality (%)			Goodness factors		Structural motifs (%)		
Rough model	Allowed	General	Disallowed	Covalent	Total	α helices	β sheets	Value of molecular pdf
Pep	99.2	0.2	0.6	−0.50	−0.29	18.0	33.5	—
Model A	93.3	3.6	3.1	−0.52	−0.32	19.8	36.4	5059.5
Model B	94.4	4.1	1.5	−0.45	−0.23	18.9	34.5	4681.8
Model C	95.6	3.0	1.5	−0.42	−0.24	18.5	34.4	4834.3

[a]Ramachandran plot qualities show the amount (%) of residues belonging to the allowed, generally allowed, and disallowed region of the plot. Goodness factors show the quality of covalent and overall bond/angle distances. These scores should be above −0.50 for a reliable model. Structural motifs (%) indicate the percentage of residues belonging to α-helix/β-sheet regions. The value of the molecular probability density function (pdf) of the different models was created by Modeller.

alignment as well as the PhD database, are depicted in Figure 2. In addition to the greatest extent of secondary structural elements identified in Model A (see Table 1) the perfect agreement between its structural motifs and those obtained by the PhD procedure allows the model to be refined.

Structures of Model A1, A2 and A3 were investigated by PROCHECK (see Table 2). The quality of the Ramachandran plot as well as the goodness factors were found to be better than those calculated for the rough Model A. The number of residues with disallowed torsional angles decreased and the quality of covalent interactions also improved. Comparing structures obtained by different refinement techniques we found that simulated annealing gave models with better quality than did constrained MD. Although this method was used for the refinement of homology models obtained for proteins even with low homology to templates, the preference of simulated annealing techniques is straightforward. Considering the fact that the conformational space can be more effectively explored by simulated annealing than simple MD, this result is not surprising at all. A comparison between Model A2 and A3 differing only in the starting temperature of the simulated annealing procedure suggests the 500 K simulation to be more effective. This result can be interpreted on the basis of side chain conformations that were dramatically altered at 900 K with respect

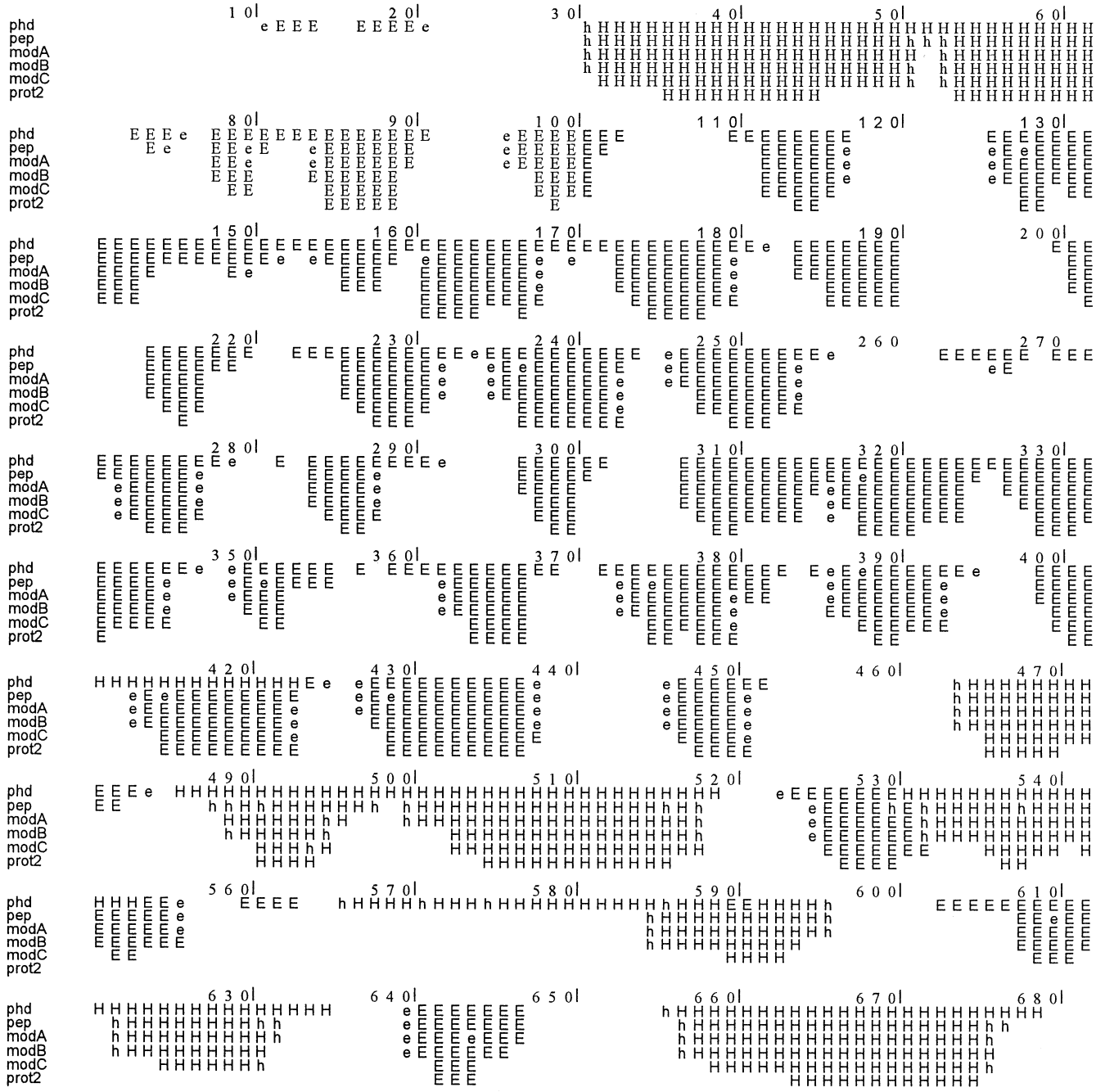


Figure 2. Comparison of secondary structures of Pep and OpB models. Secondary structure of Pep based on its X-ray structure is labelled as pep; phd stands for the secondary structure of OpB predicted by PhD; modA, modB, modC and prot2 are the secondary structures of OpB obtained from crude homology models (Model A, Model B and Model C) and the final refined model (Model A2), respectively.

to the starting ones and could not be fully optimized after the cooling period. Model A2 was finally selected as the best refined one and its quality was evaluated using three additional methods.

The first test was to compare the residue backbone conformations in Model A2 with the preferred values obtained from the Protein Data Bank of known structures. As shown in the Ramachandran plot of Figure 3, the distribution of the Ψ/Φ angles of the model is within the allowed regions. Only 14 out of the 686 (≈2%) residues have disallowed conformations. Comparing this result with that obtained for PEP, for this enzyme we found four residues located in disallowed regions. Thus, our analysis suggests the backbone conformations of Model A2 to be nearly as good as those of the template.

The second test of Model A2 was to apply energy criteria using PROSA. We investigated whether the interaction energy of each residue with the remainder of the protein is negative. PROSA energy plots for the refined Model A2 and its starting

Table 2. The Quality of Structures Obtained by Different Refinement Techniques[a]

Refined model	Ramachandran plot quality (%)			Goodness factors		rms deviation from model A		i/o distribution analysis
	Allowed	General	Disallowed	Covalent	Total	backbone	All	
A1	94.1	3.1	2.8	−0.46	−0.55	0.544	0.855	1.199
A2	95.9	2.1	2.0	0.28	−0.29	0.599	0.895	1.147
A3	95.1	2.1	2.1	0.29	−0.25	0.595	0.966	1.149

[a]The overall change in the 3D structure of OpB during different refinements: the rms deviations of the refined structures from the initial structure (model A) and the i/o distribution score of the polar/apolar amino acids (this score should be close to 1 for a well-packed water-soluble protein).

conformation (Model A) are shown in Figure 4. It can be seen that residues 350 to 400 of Model A have positive interaction energies that were eliminated during refinement. Energy analysis of the experimental structure of PEP showed no residue with positive interaction energy indicating that Model A2 fulfills the energy criterion of the PROSA analysis (Figure 5).

The third test used to evaluate our OpB model was to compare the packing environment for residues of the same type in high quality experimental structures deposited in the Protein Data Bank using the WHATIF program. A residue in a model structure with a score of −5.0 or worse usually indicates poor packing. Figure 6 shows the scores for Model A2 of OpB and for PEP as well. Since there have been no residues in Model A2 below this limit we concluded that the refined model represents an acceptable packing quality.

The final property of the OpB model investigated was its dynamic behavior during an unconstrained MD simulation. This MD test was suggested by Loew et al.[17] to check the overall stability of the model structure, since if a protein is not folded into its native state the structure will gradually denature. Accepting this rigorous test as relevant for homology models a 300 ps long unconstrained MD simulation on Model A2 was carried out. Figure 7 shows the total potential energy during the simulation. It can be seen that the system remains in equilibrium during approximately 200 ps while the protein size remains constant during the entire simulation (Figure 8). Constant radius of gyration during the entire simulation also supports the stability of the model. Fulfilling these tests, we concluded that Model A2 is stable at room temperature.

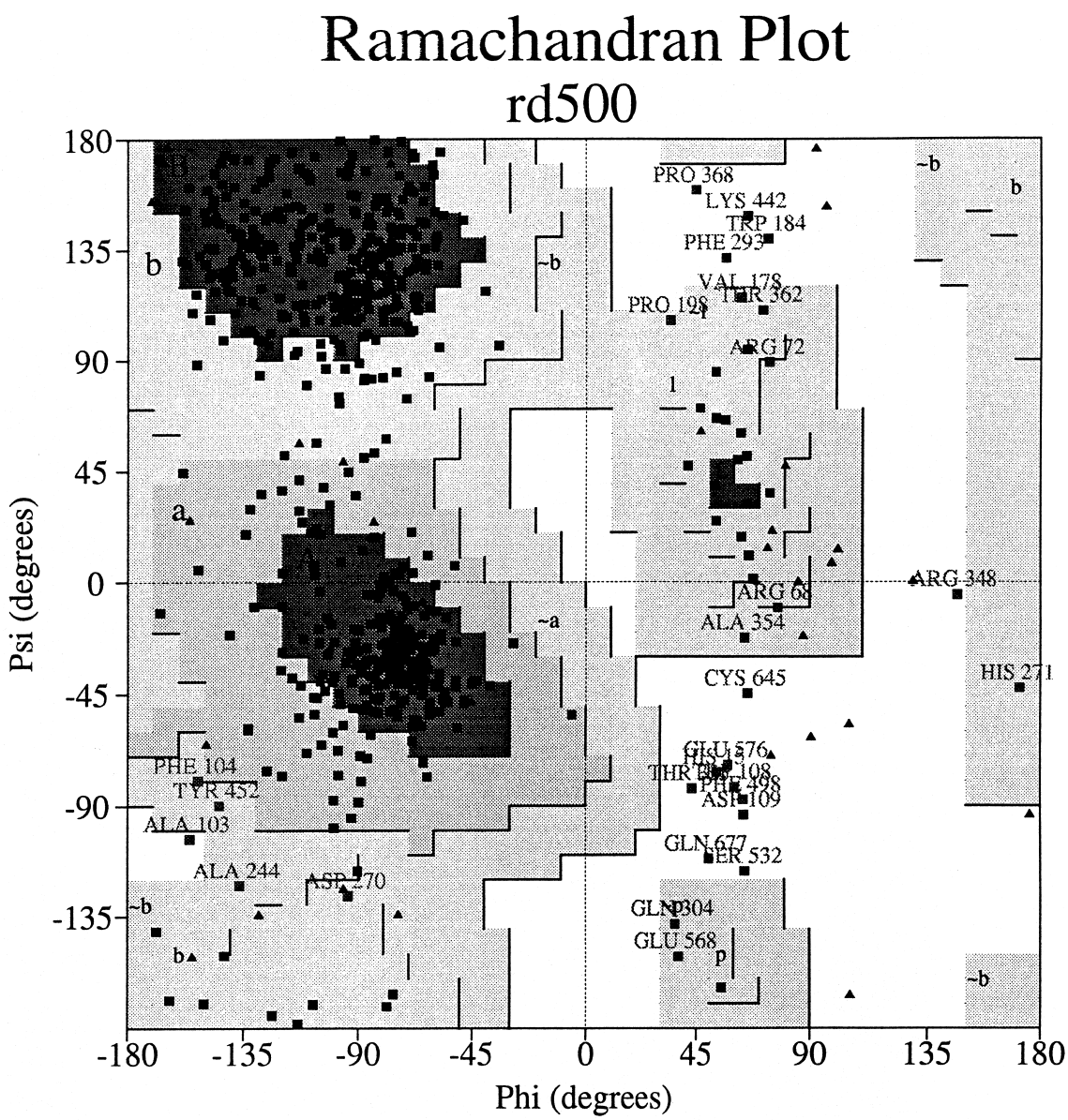
In summary, the quality of the refined model of OpB has been checked using four different types of criteria. The results showed that the backbone conformation (PROCHECK), the residue interaction (PROSA), the residue contact (WHATIF) and the dynamic stability of the structure (MD) are well within the limits established for reliable structures. Passing all tests by Model A2 suggests that we obtained an adequate model for OpB (depicted on Color plate 1) to characterize protein-substrate interactions and to investigate the relation between the structure and function.

Substrate Binding

Subjecting the secondary structure of PEP and OpB to a stricter comparison, the most notable difference is the appearance of a helical segment from Gln-587 to Lys-593 in OpB which does not appear in the secondary structure of PEP. This segment is in the vicinity of the active site in OpB (15 Å away from the catalytic Ser), in the corresponding part of the sequence of PEP is a Trp residue which is rarely part of an α-helix due to its size. A similar helical segment can be found in trypsin-type serine proteases (between residues 164 to 176 in β-trypsin).[31]

The active site (Ser-532, Asp-617, His-652, OpB numbering)[30] of the PEP type serine proteinases is located at the interface of the two domains, namely the β-propeller domain and the α/β hydrolase domain.[5] Around the substrate binding pocket of OpB, like in PEP, numerous small, flexible residues (Gly-530, Gly-531, Gly-534, Gly-535, Gly-539) can be found around the catalytic Ser avoiding steric conflicts and facilitating conformational arrangements during catalysis. There is no Cys residue in the neighborhood of the catalytic triad, which is the case in PEP since Cys-255 (PEP numbering) is deleted in OpB.

Molecular basis of the experimentally observed selectivity of OpB towards processing at dibasic sites was first investigated by the calculation of molecular electrostatic potentials within the binding site of OpB and PEP. We found that the overall binding site of OpB is more negative than the binding site of PEP (see Color Plates 2 and 3, respectively), which is in accordance with the increased activity of OpB against basic substrates. Considering the amino-acid composition of the binding site of OpB and PEP the more negative character of the OpB binding site can be rationalized by some mutation from non-polar to polar or acidic amino acid residues in OpB with respect to PEP. While Val-556 and Val-620 are conserved in the vicinity of the benzyloximino group, Tyr-455 is mutated from Phe, Thr-573 from Trp, Asp-460 and Asp-462 are mutated from Thr and Asn (OpB numbering), respectively. On the basis of Monte Carlo docking calculations we propose that Asp-462 and Asp-460 might play an important role in the binding of the Z-Arg-Arg substrate. In Pep there is a ring stacking between the indole ring of Trp-595 (Pep numbering) and the Pro residue of the substrate/inhibitor, so a Thr at this position in OpB could be more suitable for a positively charged aliphatic substrate. Interestingly there are some polar to apolar type mutations from Pep to OpB: Trp-580 is mutated from Tyr and Ala-533 is mutated from Asn. The latter one is one of the two amino acids of the oxanion hole in Pep, while Asn-555 (Pep numbering) stabilize the tetrahedral intermeder with its backbone NH. Since the secondary structure of OpB in this region remains similar to that of Pep we conlcude that the backbone NH of Ala-533 can serve a similar function than Asn in Pep.



Plot statistics

Residues in most favoured regions [A,B,L]	471	77.9%
Residues in additional allowed regions [a,b,l,p]	108	17.9%
Residues in generously allowed regions [~a,~b,~l,~p]	12	2.0%
Residues in disallowed regions	14	2.3%
	----	------
Number of non-glycine and non-proline residues	605	100.0%
Number of end-residues (excl. Gly and Pro)	298	
Number of glycine residues (shown as triangles)	45	
Number of proline residues	34	

Total number of residues	982	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Figure 3. Ramachandran plot of the ϕ/ψ distribution of Model A2 produced by PROCHECK.

Docking of Substrates to OpB

To give an explanation to the fact, that OpB catalyzes the decomposition of dibasic substrates by two orders of magnitude more effectively than monobasic ones, Z-Arg-Arg and Z-Arg (Z = benzyloximino-carbonyl) substrates were docked into the active site. Interaction energies for both substrates were also calculated.

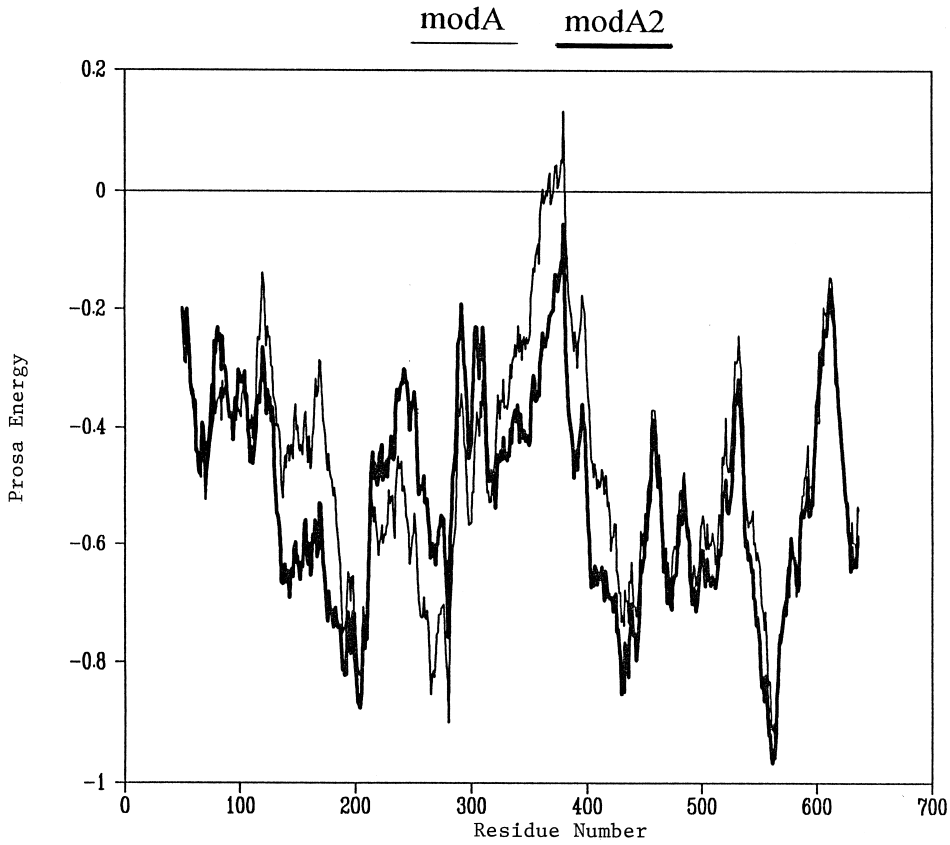


Figure 4. Prosa energy profiles calculated for the refined Model A2 and its starting conformation Model A.

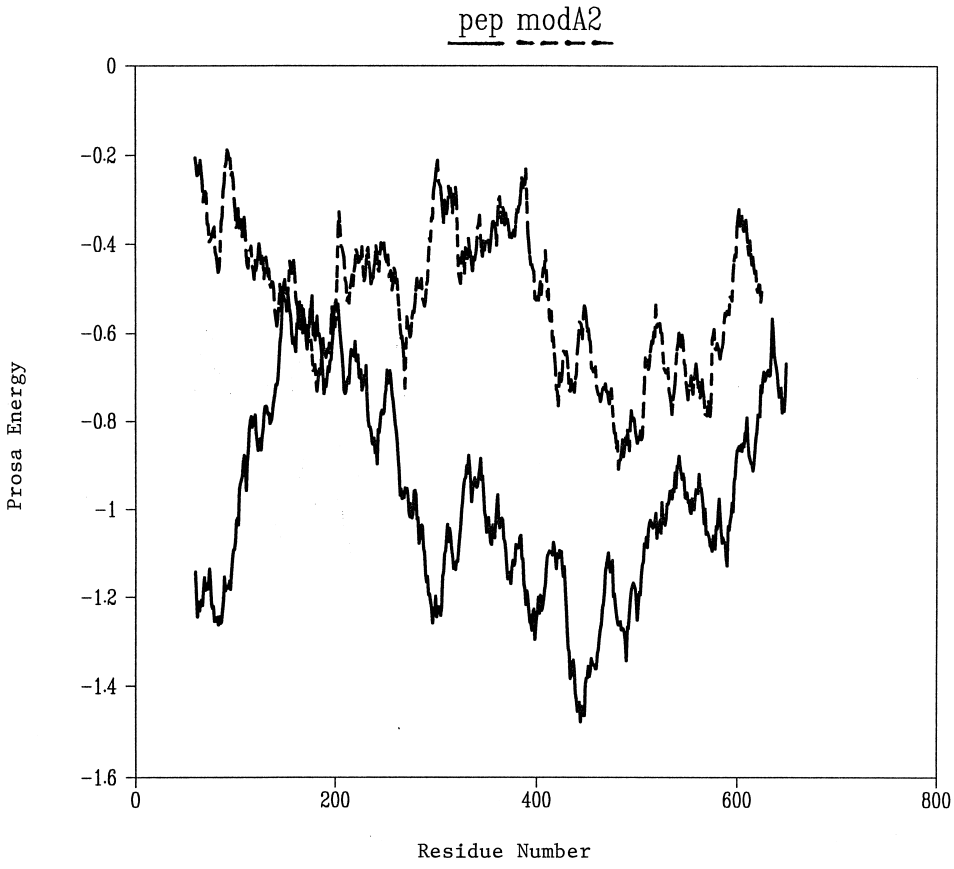


Figure 5. Prosa energy profiles calculated for the refined OpB model (Model A2) and the experimental structure of Pep.

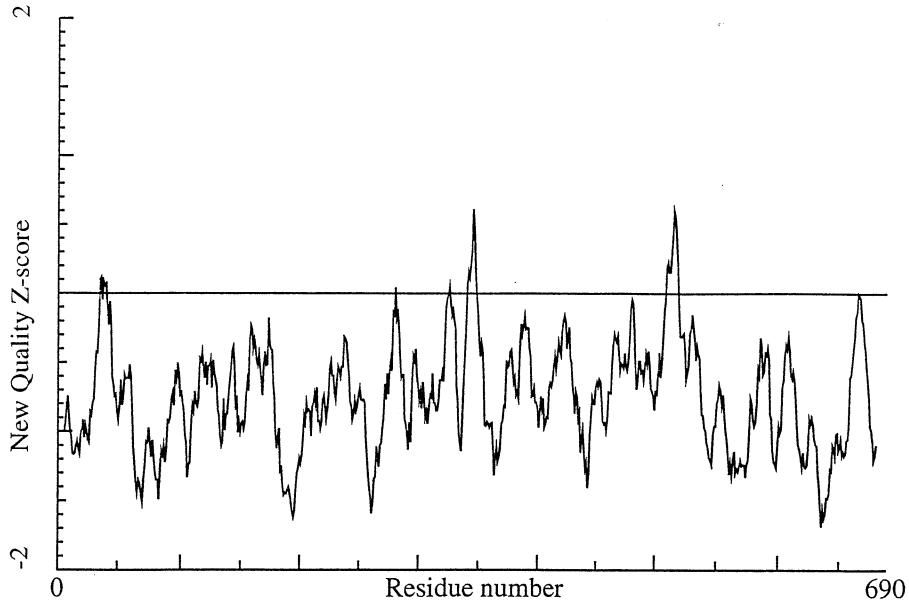


Figure 6. Whatif quality control values calculated for Model A2 of OpB.

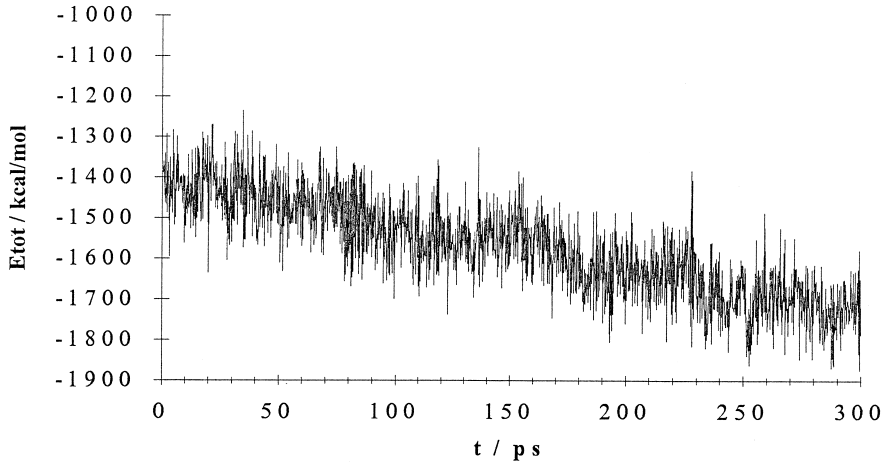


Figure 7. Total energy plot of the refined OpB model (Model A2) during 300 ps MD simulation.

Considering binding of the less effective substrate, we can conclude, that mostly nonionic residues can be found in the vicinity of Z-Arg (closer than 4 Å), namely Tyr-455, Phe-558, Ile-570, Val-620, Tyr-573, Val-556 similarly to PEP (Val-620 and Val-556 are conserved). In case of the benzyloximino group there is a tight steric fitting into the hydrophobic binding pocket. Quite a few hydrogen bonds are formed, as seen in Color plate 4, mostly with backbone oxygen atoms (Ser-532, His-652, Gly-531).

The more effective substrate could occupy a much more favorable position in the binding site. The benzyloximino group is stacked in a hydrophobic cavity (similarly to Z-Arg), but the binding of the rest of the substrate is realized mostly by polar groups, Tyr-455, Asn-619, Ala-457, Arg-333 are in the vicinity ($<$4 Å) of Z-Arg-Arg. In this case there is a possibility to form more hydrogen bonds with the hydroxyl group of Tyr-450 and Tyr-452, and with the carboxylate and backbone oxygens of Asp-460. Arg-333 makes a weak hydrogen bond to Asp-460 which makes a strong hydrogen bond to an Arg of the dibasic substrate preorganizing Asp-460 to the substrate binding.

We found an ion pair between the second Arg of the substrate and Asp-462 carboxylate oxygens (see Color plate 5). As it was first proposed by Polgár an ion pair could play an important role in the binding of the substrate and/or in the catalytic machinery, since the catalytic efficiency of OpB decreases with increasing ionic strength.[2] Thus, the above Asp side chain could be a good candidate to accomplish an ion pair with basic substrates. In the case of Z-Arg, Asp-460 is farther from the substrate, reflecting a looser binding. Calculated substrate interaction energies (-85.2 and -204.5 kJ/mol for Z-Arg and Z-Arg-Arg, respectively) are in accordance with the preferred binding of Z-Arg-Arg and supports their docked orientation to be reliable.

It is interesting to note that Tyr-450 and Tyr-452 are conserved but Asp-460 and Asp-462 are not conserved in PEP (these side chains are mutated to Thr and Asn respectively). Comparing the electrostatic fitting of both substrates (Color plates 6 and 7 for Z-Arg and Z-Arg-Arg, respectively) into the active site, we can point out, that the more effective substrate can fit more properly. Z-Arg-Arg can be fitted in a more hydrophilic region characterized by more negative electrostatic potential. In the case of Z-Arg-Arg the Arg residues are too

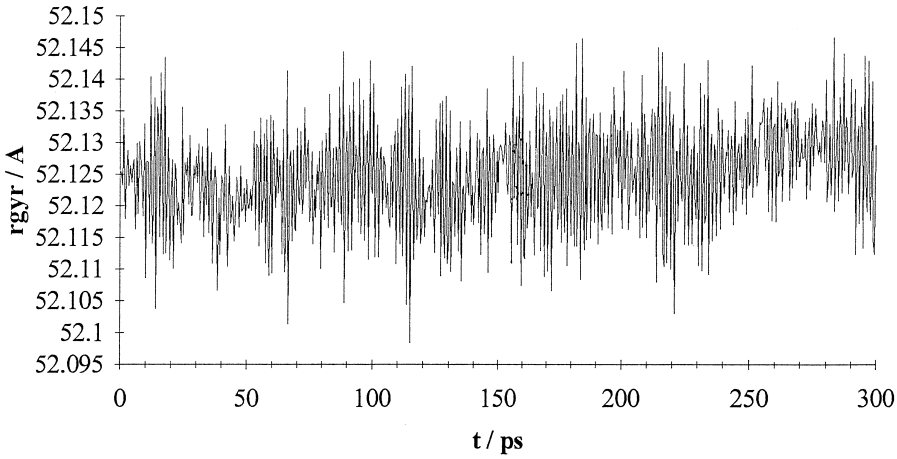


Figure 8. Radius of gyration of the refined OpB model (Model A2) during 300 ps MD simulation.

close to each other (within 4.5 Å). We tried to find analogous structures with closely packed acidic or basic residues, but these ligands were on the one hand bonded near to the solvent accessible part of the protein[32] and the other hand there were no acidic/basic residues in the vicinity of the dibasic/diacidic site giving a chance to form an ion pair.[33] In case of tyrosine kinase complexed with a fragment of the phosphotyrosine recognition domain, however, two Glu residues were 7.5 Å apart from each other with one strong ion pair. Therefore we conclude that if the active site is located in the internal region of the protein and there is a possibility to make strong hydrogen bonds and/or ion pairs with acidic residues, the dibasic substrate can occupy a tighter arrangement.

ACKNOWLEDGMENTS

This work was supported by the National Foundation for Scientific Research (OTKA F030044). The authors are grateful to Drs. Zsolt Böcskei (Budapest) and Vilmos Fülöp (Oxford) for the co-ordinates of prolyl oligopeptidase and to Prof. László Polgár (Budapest) for fruitful discussions concerning the structural interpretation of enzyme-substrate complexes.

REFERENCES

1 Steiner, D.F., Smeekens, S.P., Ohagi, S., and Chan, S.J. The new enzymology of precursor processing endoproteases. *J. Biol. Chem.* 1992, **267,** 23435–23438
2 Pacaud, M., and Richaud, C. Protease II from *Escherichia coli*: Purification and characterization. *J. Biol. Chem.* 1975, **250,** 7771–7779
3 Polgár, L. A potential processing enzyme in prokaryotes: Oligopeptidase B, a new type of serine peptidase. *Proteins: Struct. Funct. Gen.* 1997, **28,** 375–379
4 Kanatani, A., Masuda, T., Shimoda, T., Misoka, F., Xu, L.S., Yoshimoto, T., and Tsuru, D. Protease II from *Escherichia coli*: Sequencing and expression of the enzyme gene and expression of the expressed enzyme. *J. Biochem (Tokyo)* 1991, **110,** 315–320
5 Rawlings, N.D., Polgár, L., and Barrett, A.J. A new family of serine-type peptidases related to prolyl oligopeptidase. *Biochem. J.* 1991, **279,** 907–911
6 Fülöp, V., Böcskei, Z., and Polgár, L. Prolyl oligopeptidase: an unusual β-propeller domain regulates proteolysis. *Cell* 1998, **94,** 161–170
7 Böcskei, Zs., Fuxreiter, M., Náray-Szabó, G., Szabó, E., and Polgár, L. Crystallization and preliminary x-ray Analysis of porcine muscle prolyl oligopeptidase. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 1998, **D54,** 1414–1415
8 Appel, R.D., Bairoch, A., and Hochstrasser, D.F. A new generation of information retrieval tools for biologists: the example of the ExPASy WWW server. *Trends Biochem. Sci.* 1994, **19,** 258–260
9 Thompson, J.D., Higgins, D.G., and Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* 1994, **22,** 4673–4680
10 Sali, A., Blundell, T.L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 1993, **134,** 779–815
11 Sanchez, R. and Sali, A. Advances in comparative protein-structure modeling. *Curr. Opin. Struct. Biol.* 1997, **7,** 206–214
12 Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 1993, **26,** 283–291
13 Rost, B. and Sander, C. Prediction of protein secondary structure at better than 70% accuracy. *J. Mol. Biol.* 1993, **232,** 584–599
14 Brünger, A.T. X-PLOR: Version 3.1; A system for protein crystallography and NMR. New Haven, CT: Yale University Press, 1992
15 Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., and Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comp. Chem.* 1983, **4,** 187–217
16 Szilágyi, A., Závodszky, P. Structural basis for the extreme thermostability of D-glyceraldehyde-3-phosphate dehydrogenase from Thermotoga maritima: analysis based on homology modelling. *Protein. Eng.* 1995, **8,** 779–789
17 Ryckaert, J.P., Ciccotti, G., and Berendsen, H.J.C. *J. Comput. Phys.* 1977, **23,** 327
18 Sippl, M.J. Boltzmann's principle, knowledge based mean fields and protein folding. An approach to the computational determination of protein structures. *J. Comp. Aided Mol. Des.* 1993, **7,** 473–501

19 Hooft, R.W.W., Vriend, G., Sander, C., and Abola, E.E. Errors in protein structures. *Nature* 1996, **381,** 272
20 Gilson, M.K., Sharp, K., and Honig, B. Calculating the electrostatic potential of molecules in solution: method and error assessment. *J. Comp. Chem.* 1987, **9,** 327–335
21 Gilson, M.K., and Honig, B. Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies and conformational analysis. *Proteins: Struct. Funct. Gen.* 1988, **4,** 7–18
22 McDonald, D.Q. and Still, W.C. AMBER* torsional parameters for the peptide backbone. *Tetrahedron Lett.* 1992, **33,** 7743–7746
23 Mohamadi, F., Richards, N.G.J., Guida, W.C., Liskamp, R., Lipton, M., Caufield, C., Chang, G., Hendrickson, T., and Still, W.C. MacroModel-an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *J. Comput. Chem.* 1990, **11,** 440–454
24 Whitlow, M. and Teeter, M.M. *J. Am. Chem. Soc.* 1986, **108,** 7163–7172
25 Keserü, G.M., Kolossváry, I., and Bertók, B. Cytochrome P-450 catalyzed insecticide metabolism. Prediction of regio- and stereoselectivities in the primer metabolism of carbofuran: a theoretical study. *J. Am. Chem. Soc.* 1997, **119,** 5126–5131
26 Keserü G.M. and Menyhárd, D.K. The role of proximal His93 in nitric oxide binding to metmyoglobin. Application of continuum solvation in Monte Carlo protein simulations. *Biochemistry,* 1999, **38,** 6614–6622
27 Keserü, G.M., Kolossváry, I. *Molecular Mechanics and Conformational Analysis in Drug Design*, Blackwell Science, UK, 1999.
28 Tsuru, D. and Yoshimoto, T. *Meth. in Enz.* 1994, **244,** 201–215
29 Schonlein, C., Heins, J., Barth, A. Purification and characterization of prolyl endopeptidase from pig brain. *Biol Chem Hoppe Seyler* 1990, **371**(12):1159–64
30 Brendel, V., Bucher, P., Nourbakhsh, I., Blaisdell, B.E., Karlin, S. Methods and algorithms for statistical analysis of protein sequences. *Proc. Natl. Acad. Sci. USA*, 1992, **89,** 2002–2006
31 Huber, R. and Bennett, W.S. Functional significance of flexibility in proteins. *Biopolymers* 1983, **22,** 261–279
32 Feng, S., Kasahara, C., Rickles, R.J., and Schreiber, S.L. Secific interactions outside the proline rich core of SRC SH3 ligands. *Proc. Natl. Acad. Sci. U.S.A.* 1995, **92,** 12408–12415
33 Mikol, V., Baumann, G., Keller, T.H., Manning, U., and Zurini, M.G. The crystal structures of of the SH2 domain of P56LCK complexed with two phosphonopeptides suggest a gated peptide binding site. *J. Mol. Biol.* 1995, **246,** 344–355.