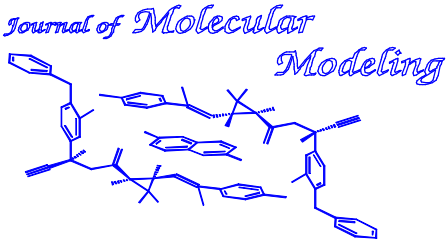
J. Mol. Model. **1995**, *1*, 54 – 67



On the Entropic and Hydrophobic Properties Involved in the Inhibitory Mechanism of Carboxypeptidase A by its Natural Inhibitor from Potato

Baldomero Oliva[1,*], Cristina Marino[1], Xavier Daura[1], Miguel A. Molina[2], Francesc Canals[1], Francesc X. Avilés[1,3] and Enrique Querol[1]

1. Institut de Biologia Fonamental. Universitat Autònoma de Barcelona. Bellaterra 08193. Barcelona. Spain (baldo@luz.uab.es)
2. Unitat de Bioquímica i Biologia Molecular. Dpt. Biologia. Facultat de Ciències Experimentals i de la Salut. Universitat de Girona. Girona 17071 Spain.
3. Dpt de Bioquímica. Universitat Autònoma de Barcelona. Bellaterra 08193. Spain.

Received: 3 January 1995 / Accepted: 14 March 1995

Abstract

The inhibition of carboxypeptidase A (CPA) by its natural inhibitor from potato (PCI) has been widely analysed by computer and experimental methods. Several mutants of PCI have been obtained in order to study the physico-chemical properties related to the inhibition. Point mutations were performed in the C-tail of PCI given its fundamental role in the inhibition. The inhibition constant and the dissociation free energy of the complexes PCI-CPA was experimentally obtained for each mutant. The mutants were divided in two sets, those where the mutation was intrinsically affecting the conformation of the PCI C-tail, and those where the mutation affected the interaction between PCI and CPA. The crystallographic structure of PCI, as found in its complex with bovine carboxypeptidase A, was used to model the structure of these mutants. Two theoretical approaches were performed to explain both sets of experimental results:
1) study of the structural features of wt PCI and mutant forms by molecular dynamics (MD) simulation, and
2) modelling of the interaction of the C-tail of PCI with CPA. The first approach provides an explanation of the observed behaviour of the first set of mutants of PCI, if the hypothesis is made of a direct relationship between the entropy of inhibition and the mobility of the C-tail of PCI. For the second set of mutants, the experimentally measured dissociation energies for the complexes PCI- CPA can be related to the theoretically estimated exposure to the solvent of the side chain of the mutated residue in the complex. In the case of the double mutation G35P+P36G, the importance of the main chain hydrogen bond between Gly 35 and Ala26, anchoring the C-tail to the core of PCI, as predicted by the MD simulations, was also corroborated by the experimental result.
The agreement between the theoretical approaches and the experimental results shows the appropriateness of our hypotheses and also the relevance of such a combined effort of experimental and computational molecular biology in protein engineering.

Keywords: Potato Carboxypeptidase Inhibitor. Molecular Dynamics, Computer Simulations, Inhibitor Engineering, Inhibition Mechanism, Excluded Volume, Free Energy of the Inhibitor-Enzyme Complex.

Abbreviations: CPA = carboxypeptidase A (EC., 3.4.17.1); PCI = carboxypeptidase inhibitor from potato; wt = recombinant wild-type; re = recombinant; del = deletion;

* *To whom correspondence should be addressed*

Introduction

Protein engineering is currently being used to analyse the structure and function of proteins and also to design proteins with specific properties that can be useful for biotechnological purposes. The main problem confronted by protein engineers is how to select amino acid replacements so that they lead to the desired properties. This is still a largely unsolved problem [1]. Point mutations can result in global changes of structure that may affect properties in an unpredictable manner. Effective application of protein engineering requires a coupling of experimental data with theoretical studies that anticipate the structural effects of a given amino acid substitution. One possible way to perform these studies is the use of Molecular Dynamics (MD) simulations [2,3] which can offer relevant information that might help to design proteins and to locate appropriate sites where mutations can be introduced [4]. Another approach is to model the interaction between proteins and ligands, obtaining the thermodynamic properties of the complexes. These approaches must be properly matched to experimental data.

Proteases and protease inhibitors are frequently subjected to protein engineering studies. These molecules are involved in important biological processes, such as hormone and neuropeptide processing, defence mechanisms, fertilisation and virus replication. It would then be very valuable to develop target-oriented inhibitors to control these processes [5]. Potato carboxypeptidase inhibitor (PCI) is a particularly suitable system for MD studies due to its small size -39 residues- and because its structure is known in aqueous solution [6], and in crystal complex with carboxypeptidase A [7] (Figure 1), thereby offering a variety of experimental data that can serve to measure the quality of the molecular simulations.

The 27-residue globular core of PCI is stabilised by three disulfide bridges and lacks any regular secondary structure, except for a short five-residue helix. Its complex with carboxypeptidase A (CPA) represents a well-known example of protein-protein interaction. A five residue C-terminal tail -residues 35 to 39- protrudes from the globular core of PCI. It docks into the active site of carboxypeptidase, leading to a stopper-like inhibition mechanism [7]. A secondary binding site, from residues 28 to 31, has also been proposed to be essential in binding [8]. The easiest-to-analyse structural/functional trait in this protein is therefore its inhibitory activity, located in the C-terminal extended tail.

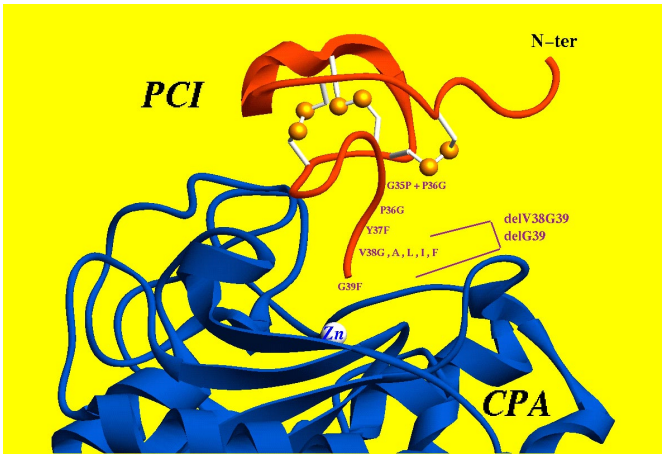


Fig. 1 *Ribbon plot of PCI interacting with CPA. The mutants of the C-tail of PCI described in the paper are indicated on the structure.*

We have previously cloned and expressed, in *E. coli*, a synthetic gene for the isoform IIa of PCI [9,10]. The recombinant protein was obtained in a soluble form, and it was shown to be identical to the PCI-IIa from potatoes. Subsequently, we obtained and analysed several mutant forms (Figure 1) which can be classified into: (a) those affecting to residues involved in the interaction with carboxypeptidase (mutants P36G, Y37F, delG39, delV38G39, V38G, V38A, V38I, V38L); and (b) those affecting the stability and functionality of PCI (mutant G35P+P36G). Additionally, the wtPCI and some of the mutant forms have been simulated and analysed by molecular dynamics (MD) [4,11,12] and others modelled on its interaction with CPA [9]. In the present article we present and discuss the rationale of an experimental/theoretical (computer simulation) approach for the study and engineering of PCI. Particular emphasis will be placed on mutants P36G, Y37F, V38G, V38A, V38I and V38L for which entropic and hydrophobic analyses were made. For P36G and Y37F the study of the conformational space of the C-tail was performed by MD analyses whilst for the rest of mutants its hydrophobic interaction was theoretically predicted from the modelling of the structure of the PCI-CPA complex. The importance of G35P+P36G mutation on the structure of PCI was also predicted by MD analyses.

Methods

Production and experimental analysis of recombinant PCI's.

Details on the construction of a synthetic gene for PCI, expression in *E coli* and purification procedures for the recombinant PCI, secreted into the culture medium, have been previously reported [10,13]. The synthetic PCI gene, and its mutant derivatives, were expressed in *E coli* MC1061 using the secretion vector pINIII-*ompA*3. Sequencing was performed in vectors M13mp18 and M13mp19, by the dideoxy method. Oligonucleotide site-directed mutagenesis was performed by the procedure of Nakayame and Eckstein [14] (Amersham kit). Wild and mutant forms of recombinant PCI, except for G35P+P36G, were purified by FPLC and HPLC, as previously described for wtPCI [10], and monitored by ELISA tests and inhibitory activity assays according to Hass and Ryan [15]. The mutants showed the same chromatographic behaviour as wtPCI, both in ion-exchange FPLC and reversed-phase HPLC, indicating that there were correctly folded [16]. The concentration of the purified solutions of mutant PCIs were quantified by UV

spectrophotometry (PCI extinction coefficient: $E_{0.1\%}$ =3.0 cm^{-1}) and also by the Bradford assay using wtPCI as a standard. The K_i values of the recombinant PCIs -wild-type and mutants- for carboxypeptidase A were calculated by Henderson method or Lineveawer Burk procedure [17]: hyppuril-L-phenylalanine was used as a substrate at different concentrations, and PCI concentration was varied for each substrate concentration. The enzyme was bovine carboxypeptidase A at 42.5 nM in all cases, and UV-absorption measurements at 254 nm were made every 20 sec over 2 min. Velocities were expressed as the slope of the linear increase in absorbance. These data were used to calculate K_i. Several independent K_i measurements were made for wild-type and mutant rePCIs, the average K_i being derived from these data. ΔG_d^o values were calculated according to the equation $\Delta G_d^o = -RTlnK_i$. .

The mutant PCI G35P+P36G was impossible to obtain. It could not be detected in cultures of *E. coli* carrying the appropriate gene. This fact suggests that this mutant can not fold properly and therefore is quickly degraded.

Molecular dynamics analysis.

It was performed to analyse mutants P36G and Y37F. Molecular dynamics calculations were carried out on a Personal Iris 4D/35, on a Crimson Elan (Silicon Graphics) and on a CRAY X-MP. The X-ray crystallographic coordinates of PCI-IIa isoform in complex with CPA were used to seed the MD simulation of the wtPCI after removal of the N-terminal residue Glu1 (some isoforms lack it) and the C-terminal residue Gly39, which is not covalently bound to PCI in its complex with CPA. The same structure, after the substitution of Pro36 by Gly and Tyr37 by Phe, was used to seed the MD simulations of the mutant PCIs. The structural patterns were derived with the GROMOS program package [18]. The standard GROMOS NIS (D4) potential energy function was used for the simulations.

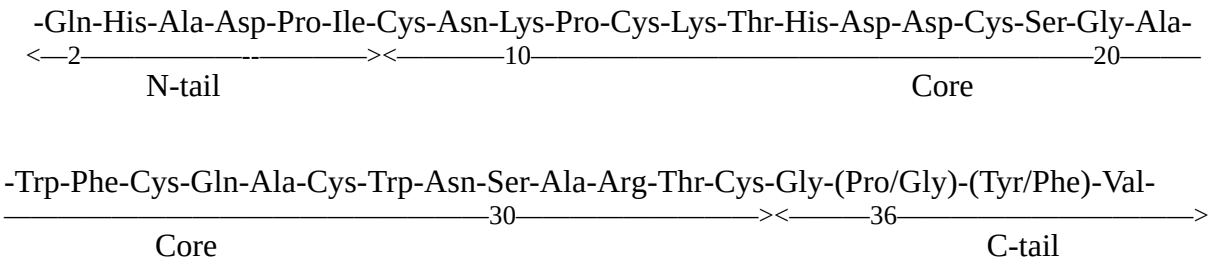
The cut-off for the energy potential was 8 Å, without use of a switching function. A cut-off of 13 Å was used for the long-range interactions. A 2 fs time step was used in the integration, and the bond lengths were constrained by SHAKE [19]. The temperature here analysed was set at 293° K, and a temperature relaxation time of 0.1 ps was used to maintain the system in a weak coupling bath. A 1800 step of steepest descent energy optimization was carried out with each structure; bond lengths were not constrained in the optimization. This energy optimization was done with the only purpose of quickly provide an initial structure to seed the MD simulations. The convergence criterion was a difference between the penultimate and the last energy smaller than $1x10^{-3}$ $kJ \cdot mol^{-1}$.

For MD simulations the structures of PCI, wild-type and mutants, were divided into three domains as indicated in Scheme 1: the core, N-tail and C-tail. The core is defined from Cys8 to Cys34, the N-tail from Gln2 to Ile7 and the C-tail from Gly35 to Val38. In this way the core contains the three disulfide bridges and both tails begin with a residue followed by a Pro in the terminal direction.

The conformational space of the C-tail for wtPCI, P36G and Y37F was studied. The analysis of changes in orientation of the PCI C-tail involves the definition of two angles: Ψ, θ and a plane. They are depicted in Figure 2. The plane is fixed in three points provided by the atomic average positions, in the crystallographic structure, of Phe23, Ala26, and Asn29. Thereafter, the normal (N) is defined. The residues Gly35, Pro36 (Gly36 in one mutant) and Tyr37 (Phe37 in the other mutant) define the first part of the C-tail. The vector CT1 is defined as going from the α–carbon of Gly35 to the α–carbon of Tyr37 (or Phe37). The residues Tyr37 (or Phe 37) and Val38 define the second part of the C-tail. The vector CT2 is defined as going from the α–carbon of Tyr37 (or Phe 37) to the α–carbon of Val38. The angle Ψ is that between the vectors N and CT1. This angle describes orientations of the first part in the C-tail. In the second part of the C-tail the vectors N and CT2 define the angle θ. We can calculate the probability density function for each angle, δP(Q) from the histogram of the angles sampled by MD:

$$dP(Q) = 1/ Ntotal \cdot dQ,QMD(tn)$$

where N_{total} is the total number of conformations sampled in the trajectory; $Q_{MD(tn)}$ is the angle Q sampled at time tn; and δQ, Q_{MD} is a function that is set equal to one if the difference



Scheme 1

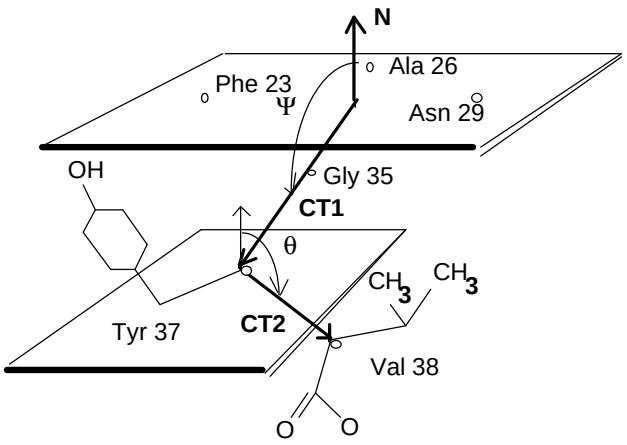


Fig. 2 *Geometrical definitions for the set of angles characterising the C-tail orientation of PCI.*

between Q and $Q_{MD(tn)}$ is positive and smaller than one, zero otherwise.

Modelling of the mutant structures on residue 38.

The structure of wild type PCI was taken from the X-ray structure of the PCI-IIa isoform in the complex with CPA [7], adding the C-terminal residue Gly. The structure of the PCI mutants and the corresponding complexes with CPA were modelled and visualised from the wtPCI structure by means of TURBO FRODO program [20]. These structures were optimised by 5000 steps of steepest descent using the GROMOS package under NIS force field [21]. The convergence criterion for the minimisation was a difference between the penultimate and last energy smaller than $1x10^{-6}$ kJ·mol^{-1}. The positions of the main chain atoms of PCI were constrained, the mutant side chains were built over them, and an optimization of the structure by energy minimisation was performed to avoid non-allowed contacts of the mutated side chains. This approach assumes that the conformational changes in mutant PCIs are small and do not significantly perturb positions of the main chain atoms of PCI in PCI-CPA complex. We think this approach to be valid by two reasons:

(1) First of all, because the inhibitor establishes many contacts with CPA [7] and only a few of them are directly affected by the mutations of Val38.

(2) More important yet, because all the mutants keep a considerable inhibitory power (see results), particularly those carrying the bulkiest residues. The calculation of the volumes of the optimised structures were made by GEPOL[22].

A hypothesis about the hydrophobic effect on the inhibition due to point mutations on residue 38 was developed as follows. Assuming that the differences between the interaction of PCI with CPA and the interaction of any of the selected mutant types of PCI (V38G, V38A, V38L, and V38I) with CPA, are mainly due to the hydrophobic effects of the replacement in residue 38 of PCI, we may consider that the binding of the inhibitor in vacuum for the wild type of PCI and for these mutant types of PCI have the same free energy exchange,

$$\Delta G_{wt}^{I,v} = \Delta G_{m}^{I,v}$$

but not in solvent, therefore,

$$\Delta G_{wt}^{I,s} \neq \Delta G_{m}^{I,s}$$

being *wt* for wild-type, *m* for mutant, *I,v* for inhibition in vacuum and *I,s* for inhibition in solvent (water).

It can be derived that :

$$\Delta G_{m}^{I,s} - \Delta G_{wt}^{I,s} = \left(\Delta G^{s}_{m,(CPA-PCI)} - \Delta G^{s}_{wt,(CPA-PCI)}\right) - \left(\Delta G^{s}_{m} - \Delta G^{s}_{wt}\right) \quad (1)$$

being *s* for solvation energy, *m,(CPA-PCI)* for mutant-type PCI in complex with CPA and *wt,(CPA-PCI)* for wild-type PCI in complex with CPA.

From Richmond [23], and considering that the free energy exchange is proportional to the excluded volume (i.e., the increase of water molecules in the environment excluded from the outer shell of solvent of the PCI and the CPA by the inhibition), we may write:

$$\Delta G^{s}_{m,(CPA-PCI)} - \Delta G^{s}_{CPA} - \Delta G^{s}_{m} = -k\,(V_{m}^{excl}) \quad (2)$$

$$\Delta G^{s}_{wt,(CPA-PCI)} - \Delta G^{s}_{CPA} - \Delta G^{s}_{wt} = -k\,(V_{wt}^{excl}) \quad (3)$$

where k is a constant ($k = b^{-1}r$, with $b=K_bT$ and r the molecular density of the excluded solvent) and

$$V_{m}^{excl} = V_{CPA} + V_{mPCI} - V_{m,(CPA-PCI)} \quad (4)$$

$$V_{wt}^{excl} = V_{CPA} + V_{wtPCI} - V_{wt\,(CPA-PCI)} \quad (5)$$

being V^{excl} the excluded volume for each structure defined in the subscripts.

Therefore, by direct substitution of (4) and (5) into (2) and (3), and this into (1) it can be obtained:

$$\Delta G^{s}_{m,(CPA-PCI)} - \Delta G^{s}_{wt,(CPA-PCI)} = -k\left(V_{m}^{excl} - V_{wt}^{excl}\right) = -k\Delta V^{excl} \quad (6)$$

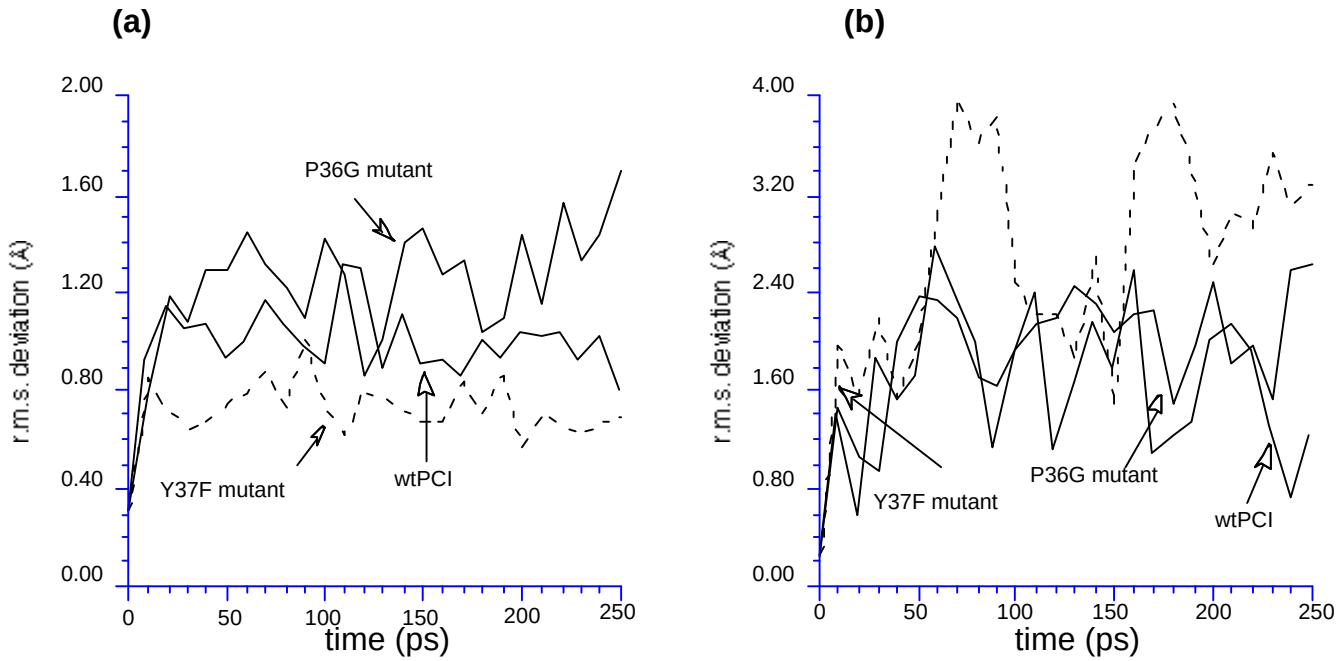


Fig. 3 *r.m.s. deviation of the MD equilibrated structure with respect to the X-ray structure, for the a-carbons. (a) wt and mutant PCI-cores; (b) wt and mutant PCI-C-tails*

The calculation of ΔV^{excl} can be directly obtained from the volume of each species that is calculated with GEPOL [22]. This hypothesis provides therefore a verifiable relationship between experiment and theory.

Results and discussion

A.- Entropic contribution of the C-tail on the inhibition and interactions responsible for C-tail orientation.

1.- Site-directed mutagenesis studies of the mutant forms P36G and Y37F.

We substituted Pro36 for Gly to test the role of these residue in the orientation of the PCI C-tail. In spite of the fact that glycine can be considered the more drastic replacement in position 36, since it should lead to an increased freedom of movement in the PCI tail, we should note that is not uncommon to find substitution of a proline by a glycine in homologous sequences from different species [24,25]. In addition, the mutant Y37F was obtained to prevent the hypothetical anchoring of the C-tail to the core of PCI by hydrogen bonding of the hydroxyl group of Tyr37 with His16, previously suggested by our MD studies (see Table 2). Prior to the theoretical and experimental analysis we considered that, specially in the case of the P36G mutant, the inhibitory activity should be dramatically affected due to higher freedom of movement in the mutant PCI C-tail. It could be expected that this enhanced freedom would decrease the probability of fitting between PCI and CPA.

Mutants P36G and Y37F were created by site-directed mutagenesis, the recombinant PCIs expressed and their inhibitory activities assayed as described in Methods.

The inhibition constant, K_i, of competitive inhibitors corresponds to the dissociation constant of the enzyme-inhibitor complex [26], K_d. Calculation of the dissociation constants of competitive inhibitors from kinetic studies are valid though the K_M of the enzyme-substrate system does not represent a true dissociation constant [27]. The standard free energy of dissociation, ΔG_d^o, for the inhibitor-enzyme complex can be calculated as:

$$ÆG_d^o = -RT\ln K_d$$

Accordingly, the standard association free energy of the inhibitor-enzyme complex will be $-\Delta G_d^o$.

Several independent K_i measurements were obtained for wild-type PCI and mutants P36G and Y37F. The average K_i and the derived Gibbs energies of dissociation, ΔG_d^o, were calculated from these data as shown in Table 1 . Surprisingly, both mutants showed a K_i indistinguishable from that of wtPCI

2.- Computer simulations: wtPCI, P36G and Y37F Structural Patterns.

1.- Structural features. The three systems: wtPCI, P36G and Y37F, reach its energetic and structural equilibrium points after about 50 ps of simulation. The average potential energy in this equilibrium was found between -1140 $kJ \cdot mol^{-1}$ and -1180 $kJ \cdot mol^{-1}$ in all cases. Core α–carbons r.m.s. deviation time series are depicted in Figure 3a. The r.m.s deviate less than 1.5 Å in the three cases for the MD-equili-

Complex	K_d (Interval)	K_d Mean	-ΔG interval (kcal·mol^{-1})	-ΔG Mean (kcal·mol^{-1})
rewtPCI-CPA	0.9-2.1E-9	1.5E-9	11.6-12.1	11.8
rePCI P36G-CPA	1.0-2.7E-9	1.9E-9	11.4-12.0	11.7
rePCI Y37F-CPA	0.3-2.8E-9	1.2E-9	11.5-12.8	12.0
rePCIdelG39-CPA	3.1-4.3E-9	3.7E-9	11.3-11.5	11.4
rePCIV38A-CPA	65-95E-9	76E-9	9.4-9.6	9.5
rePCIV38G-CPA	0.74-0.9E-6	0.82E-6	8.1-8.2	8.1
rePCIdelV38G39-CPA	26-55E-6	41E-6	5.7-6.1	5.9
rePCIV38L-CPA	11.4-14.7E-9	13.0E-9	10.5-10.7	10.6
rePCIV38I-CPA	1.7-2.5E-9	2.1E-9	11.5-11.7	11.6

Table 1. *Gibbs energy (ΔG) of the dissociation of the complexes PCI-CPA*

brated structures versus the X-ray wild-type structure. However, the MD-PCI P36G mutant structure deviates from the crystallographic wild-type structure more than the MD-wild type, especially after 180 ps. In the case of the Y37F mutant the core structure is shown to be more stable than the wild-type and it deviates less than 0.8 Å from the latter structure, showing both a similar conformational space and dynamic properties.

To calculate the C-tail r.m.s. deviations, the core Cα atoms were fitted to the crystallographic structure. The C-tail r.m.s. deviations of wild type and mutants oscillate between 0.5 Å and 2 Å. (Figure 3b). In spite of relatively large fluctuations that are apparent here, the C-tail wanders in a region not too far from the one found in the X-ray structure in all three cases. It can be concluded that PCI has a C-tail conformation space favouring formation of a complex with the enzyme. Moreover, when the local RMSD of the C-tail conformation is calculated by fitting to the crystallographic structure the Ca atoms of residues 33 to 38 of PCI, a deviation of about 0.5 Å for the wild and mutant types averaged structures respect to the X-ray structure is found. This shows that the local conformation of the C-tail has been maintained during the simulation. The RMSD of the wild type MD averaged structure of PCI was about 0.45 Å respect to the complexed structure.

2.- Hydrogen bonds. Table 2 gives a list of hydrogen bonds established with a probability higher than 40% obtained from the 50 to 250 ps MD-simulations. Table 2 shows that the hydrogen bond pattern of the wtPCI conformational space along the MD simulation and the Y37F mutant are fairly similar. The table also shows the hydrogen bond pattern of the crystallographic structure, indicated with the symbol "+". Figure 4 depicts the net of hydrogen bonds established around the elbow of the C-tail in wtPCI.

With regard to the C-tail behaviour, several conclusions can be drawn from the hydrogen bond percentages. The mainchain H-bond Gly35-Ala26 (this hydrogen bond fixes the first part of the C-tail during the time span between 50ps until 250ps) appears in the X-ray structure, both mutants and wild-type MD-structures. This H-bond does not explain why the C-tail does not show much larger deviations than the core. A similar effect is produced by the mainchain hydrogen bond Tyr37-Phe23, which fixes the last part of the C-tail, although the strength of this H-bond is decreased in the P36G mutant and increased in the Y37F mutant. The negligible importance of the H-bond of the phenolic hydroxyl of Tyr 37 with the amine on the histidine ring of His 15 is remarkable , this hydrogen bond disappears on the MD simulation of the wild type PCI.

3.- Approximation to the entropic properties of the inhibitory mechanism: Distribution of the angles defining the C-tail. The study of the C-tail angles (Figure 5 a,b) aims at a more quantitative characterisation of this primary binding zone. For the

Donor atom				Acceptor atom		X-RAY occurrence	PCI occurrence %	P36G occurrence %	Y37F occurrence %
Mainchain - Mainchain hydrogen bonds									
5	ASP	N	3	HIS	O	-	85.2	84.9	82.8
7	ILE	N	20	GLY	O	-	75.9	-	-
8	CYS	N	5	ASP	O	+	-	-	-
9	ASN	N	34	CYS	O	+	71.8	48.9	77.7
10	LYS	N	7	I LE	O	+	-	-	-
10	LYS	N	8	CYS	O	-	76.6	58.6	60.9
12	CYS	N	32	ARG	O	+	-	84.5	57.0
17	ASP	N	14	THR	O	-	95.2	76.2	96.0
18	CYS	N	15	HIS	O	-	77.1	54.7	78.6
21	ALA	N	19	SER	O	-	49.6	-	-
23	PHE	N	21	ALA	O	+	-	-	-
24	CYS	N	21	ALA	O	-	88.4	73.4	81.7
25	GLN	N	22	TRP	O	-	49.0	-	49.5
26	ALA	N	35	GLY	O	+	-	55.2	-
28	TRP	N	33	THR	O	-	64.6	-	48.1
30	SER	N	28	TRP	O	+	-	-	-
33	THR	N	31	ALA	O	-	-	50.9	-
34	CYS	N	10	LYS	O	+	63.5	85.8	66.4
35	GLY	N	26	ALA	O	+	76.7	50.5	79.0
37	TYR	N	23	PHE	O	-	71.1	66.6	92.0

Table 2a. *Hydrogen bond statistics for wtPCI and mutants P36G and Y37F between 50 to 250 ps. Mainchain- Mainchain hydrogen bonds*

X-ray structure the following values are found: Ψ. = 153.3°; θ. = 131.2°.

Angle Ψ. The probability density distribution for Ψ in the wild type and P36G, Y37F mutants has a gaussian form. The main maximum spans from 90° to 135° for wild-type PCI, 90° to 145° for the P36G mutant, and 80° to 120° for the Y37F mutant. The tails of both mutants spend most of the time in a region of conformational space which corresponds to that of wtPCI (which roughly corresponds to that found in the crystallographic orientation). In summary, there is a good match in the conformational-space probability area of the angle Ψ between the wt PCI and the mutants.

Angle θ. This probability function shows a broad distribution in the wild-type and both mutants. At angles close to the crystallographic value, the function has significant values for all simulations. The highest maximum for the wild-type is around 135°, while for the mutant P36G it is around 110° and 150° in the case of the Y37F mutant. The density of the probability function shows that the end of the C-tail remains in both mutants sticking out of the protein core, as happens in the wild-type.

The conclusion is that mutants P36G and Y37F will inhibit CPA, but it cannot be easily determined to what extent the inhibition will be affected. The observed behaviour of the C-tail for wtPCI, P36G and Y37F shows a high similarity in conformational space (see figure 5) defined by the distribution of the angles Ψ and θ, as well as by the r.m.s. (see Figure 3b). This similarity can be regarded in thermodynamic terms as entropy and we may then assume that the

Donor atom			Acceptor atom			X-RAY occurrence	PCI occurrence %	P36G occurrence %	Y37F occurrence %
Mainchain - Sidechain hydrogen bonds									
7	ILE	N	5	ASP	OD1	+	-	-	47.1
14	THR	OG1	13	LYS	O	-	-	91.6	-
14	THR	N	17	ASP	OD1	-	42.4	-	44.1
15	HIS	N	15	HIS	ND1	-	55.7	40.0	52.6
21	ALA	N	25	GLN	OE1	-	-	-	56.3
22	TRP	NE1	4	ALA	O	+	-	-	-
22	TRP	N	25	GLN	OE1	-	80.7	-	-
23	PHE	N	5	ASP	OD2	+	-	-	-
25	GLN	NE2	18	CYS	O	-	-	-	83.9
25	GLN	NE2	22	TRP	O	+	-	-	-
33	THR	OG1	31	ALA	O	-	94.6	-	92.9
37	TYR	OH	22	TRP	O	-	-	40.7	-
Sidechain - Sidechain hydrogen bonds									
9	ASN	ND2	3	HIS	NE2	-	61.3	-	51.0
15	HIS	ND1	37	TYR	OH	+	-	-	-

Table 2b. *Hydrogen bond statistics for wtPCI and mutants P36G and Y37F between 50 to 250 ps. Sidechain hydrogen bonds*

variation on the entropy for the inhibition of CPA has small differences between wtPCI and both mutant form P36G and Y37F. In addition, the interaction between PCI and CPA would be similar for wtPCI, P36G and Y37F, because the proline residue in position 36 has few contacts with CPA and the benzene ring of residue 37 is maintained in Y37F. Although the hydrogen bond of the phenolic hydroxyl of residue 37 is lost on the Y37F mutant, this does not affect to the C-tail conformation. In conclusion both mutants would have an inhibition constant similar to the one for wtPCI.

From all these studies, it can be concluded that the orientation of the C-tail of PCI cannot be attributed to the directionality imposed by proline 36, being the key factor for the C-tail orientation the net of hydrogen bonds between the whole C-tail and the PCI core (see figure 4) . This results stem the hypothesis that the mainchain H-bond Gly35-Ala26 may be of crucial importance on the inhibitory function of PCI. In order to further test this hypothesis, the mutant gene G35P+P36G was obtained. The substitution of Gly35 by Pro would suppress this hydrogen bond. Moreover, changing Pro36 by Gly would help to retain the C-tail conformation present in the PCI native form. When we tried to express the mutant gene G35P+P36G, no mutant protein was found neither by activity assays nor by ELISA, suggesting that this mutant can not fold properly and it is quickly degraded. This surprising result shows that the net of hydrogen bonds connecting the core and the C-tail is probably involved not only in the ability of PCI to inhibit CPA but also in the folding pathway of PCI itself.

B.- Hydrophobic contribution of residue 38 on the inhibition.

1.- Site directed mutagenesis studies of the mutant forms: delG39, delV38G39, V38G, V38A, V38L and V38I.

To test the role of Val 38 in the stability of the PCI-CPA complex, a series of site-directed mutants was created: delG39, delV38G39, V38G, V38A, V38L and V38I. All of these mutants were expressed in *E. coli* as extracellular soluble proteins. Their *K*i and ΔG is shown in Table 1.

The K_i of the first mutant of the series, delG39, was 3.1-4.3 x 10^{-9} M, only slightly higher than that of wild-type PCI, 0.9-2.1 x 10^{-9} M, indicating that the last residue of the C-tail of PCI, Gly 39, has no important role in the stabilisation of its

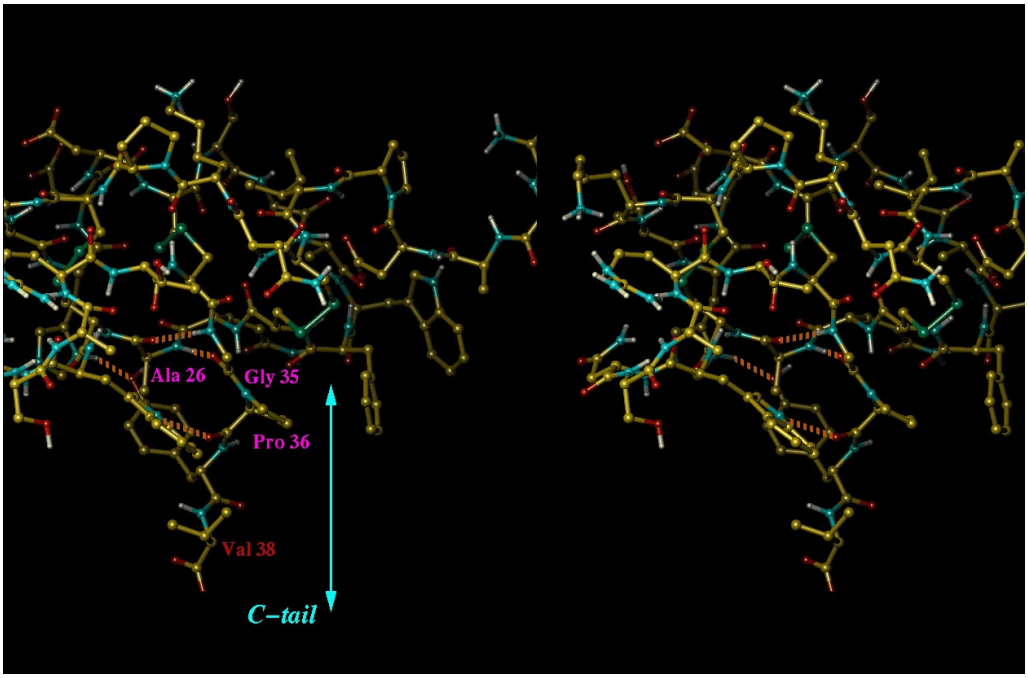


Fig. 4 *Hydrogen bonds net anchoring the C-tail of wtPCI according to the crystallographic structure.*

complex with CPA. In fact, a highly homologous inhibitor from tomato lacks the C-terminal glycine [28]. In contrast, the mutant PCI with deletion of the last two residues of the C-tail, Val38 and Gly39 (PCI delV38G39), showed a dramatic decrease of inhibitory activity towards CPA. The stability of the complex PCIdelV38G39-CPA, expressed as ΔG_d^o, decreased to 5.7-6.1 kcal·mol^{-1} (see figure 6). The overall contribution of Val38 to the stability of the complex PCI-CPA can be estimated by comparing the ΔG_d^o obtained for the complex PCIdelV38G39-CPA with that of the complex PCIdelG39-CPA. We can therefore estimate that Val38 contributes with 5.2-5.8 kcal·mol^{-1} to the stability of the natural complex. This value represents about half of the total stability of the complex, 11.6-12.1 kcal·mol^{-1}.

The point mutants V38A and V38G were obtained to estimate the contribution of the interactions established by the

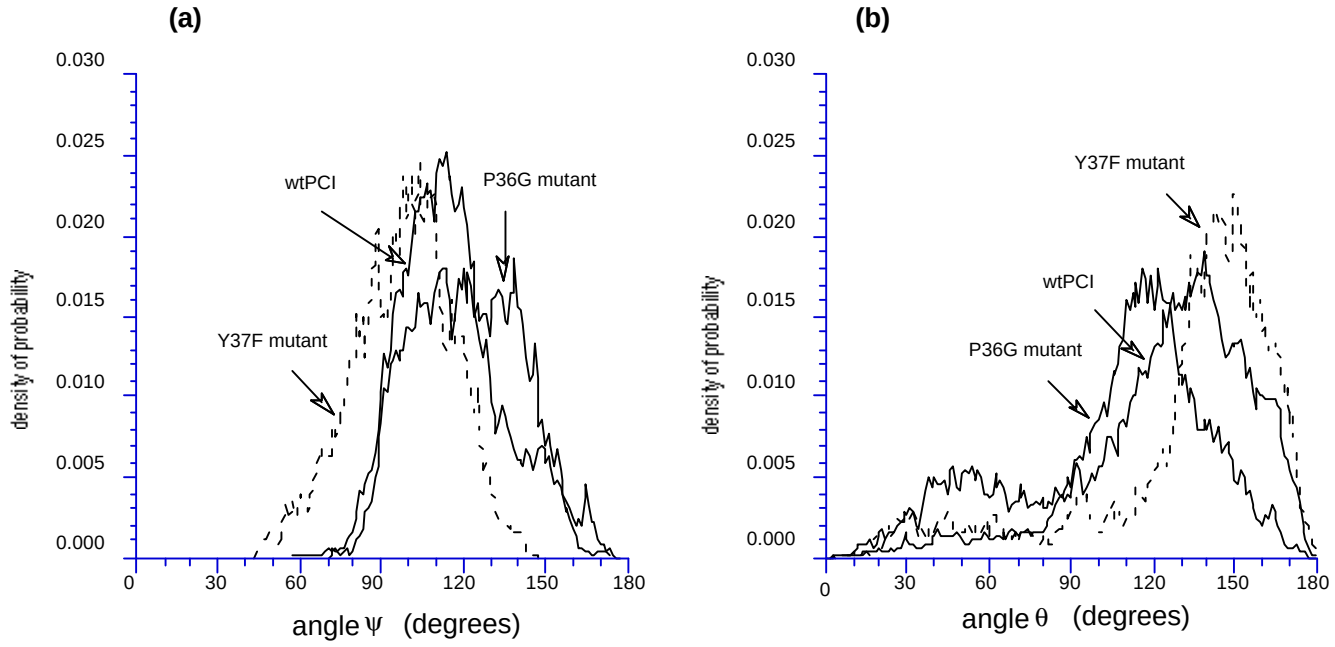


Fig. 5 *Angular distribution function for the angles defined to characterise the C-tail orientation of PCI.*

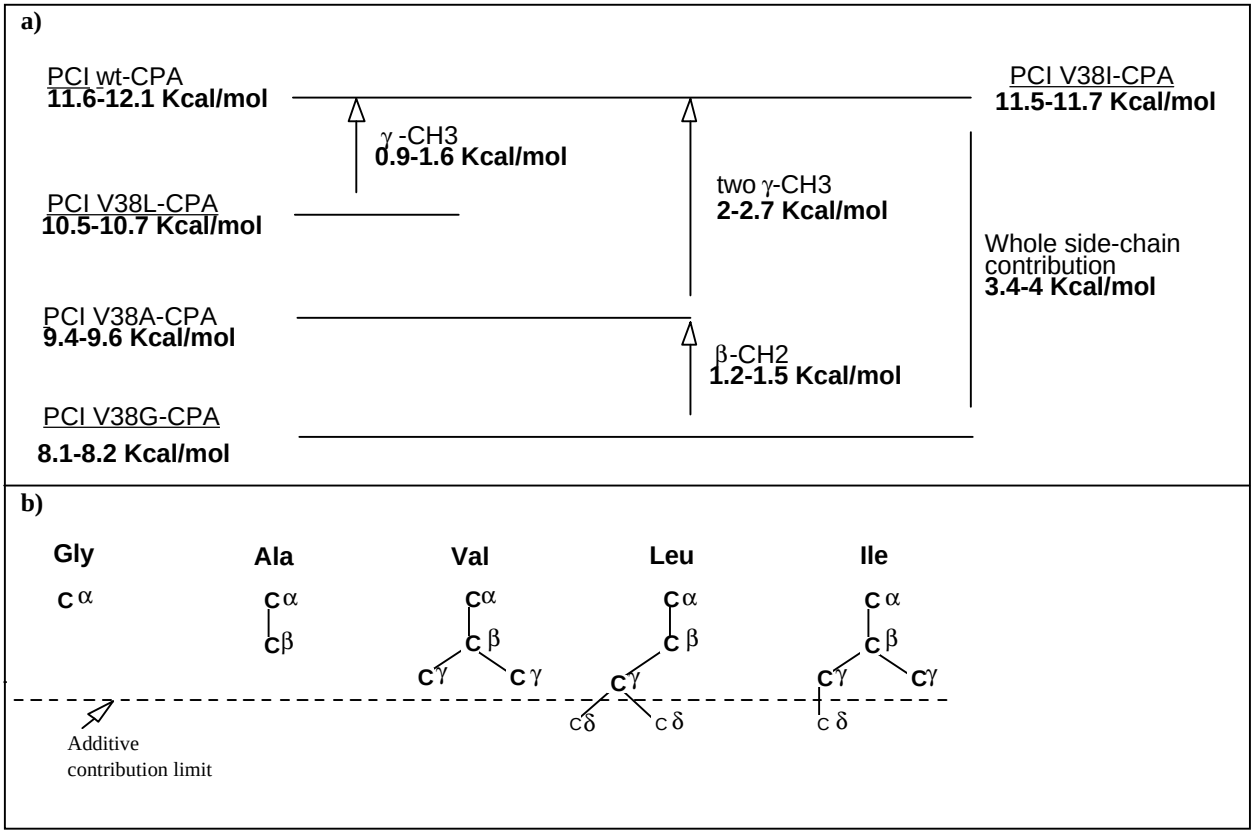


Fig. 6 *(a) Differences in dissociation free energies (ΔG_d^o) of the complexes formed by CPA and the recombinant wtPCI, V38L, V38I, V38A,V38G. (b) Schematic representation of the side chain groups of the residue in position 38 in wtPCI, V38G, V38A, V38L, and V38I. Methyl groups in position δ, present in mutants V38L and V38I, do not contribute to the stability of the PCI-CPA complex.*

different chemical groups of PCI Val38 with residues at the CPA active site. The difference in ΔG_d^o between the complexes of CPA with PCI V38G and with wt PCI, 3.4-4 kcal·mol^{-1}, is an estimation of the hydrophobic contribution of the three aliphatic side chain carbons to the overall stability of the complex (Figure 6a). The difference between this contribution (3.4-4 kcal·mol^{-1}) and the overall contribution of Val38 (5.2-5.8 kcal·mol^{-1}) is 1.2-2.4 kcal·mol^{-1}, which can be attributed to the interactions that the carboxyl group of Val38 establishes with CPA. The second mutant, PCI V38A, showed a free energy of dissociation with CPA of 9.4-9.6 kcal·mol^{-1}. The difference in K_i and stability of the complex PCI-CPA between both mutants (V38A and V38G) is, therefore, attributable to the hydrophobic contribution of the side chain methyl group in position 38 of PCI V38A. This contribution can be estimated as 1.2-1.5 kcal·mol^{-1}. It can be assumed that the contribution of the methylene in position b of Val38 to the overall stability of the complex wt PCI-CPA is then 1.2-1.5 kcal·mol^{-1}. The difference in DG_d^o between the complexes of wtPCI-CPA and PCI V38A-CPA is 2-2.7 kcal·mol^{-1}, attributable to the two g-methyl groups of Val38, absent in the mutant PCI V38A (Figure 6a,b). As both groups are sterically equivalent, they are likely to contribute equally, 1-1.4 kcal·mol^{-1}, to the overall stability of the PCI-CPA complex. This value is similar to the contribution of the b-methylene of Val38 deduced above. These results suggest that the hydrophobic contribution of each methyl/methylene group of residue 38 to the ΔG_d^o of the PCI-CPA complex is the same, about 1-1.5 kcal·mol^{-1}.

The mutants so far analysed indicated that the side-chain hydrophobic contribution of the residue in position 38 of PCI is very important for the stability of the protease-inhibitor complex. In order to further analyse it we constructed two more mutants: V38L, V38I. The mutant V38L was less active than wtPCI (the free energy of the complex V38L-CPA is 10.5-10.7 kcal·mol^{-1}). In contrast, the mutant V38I presented a K_i indistinguishable from that of wt PCI. In order to explain these results, we formulated the hypothesis that the stability of the complex PCI-CPA depends not on the overall side chain hydrophobicity of the residue 38 of PCI but only on the additive hydrophobic contributions of the methylene/methyl groups in positions β and γ (Figure 6b). The aliphatic groups in position δ of the residue 38 of PCI, present in mutants V38L and V38I, would not make any contribution to the stability of PCI-CPA complex. This hypothesis is supported by the results obtained of ΔG_d^o for V38L and V38I. The CPA complex with

V38L should have a ΔG_d^o 1-1.5 kcal·mol^{-1} lower than the complex with wt PCI, since it loses one of the methylene groups in g that contributed 1-1.5 kcal·mol^{-1} to the binding. The difference found experimentally between the ΔG_d^o of the two complexes was 0.9-1.6 kcal·mol^{-1} (figure 6a). In the other hand, valine and isoleucine have the same number of methyl/methylene groups in β and γ positions of their side chains. Therefore, their complexes with CPA should have the same ΔG_d^o, and that is exactly what was experimentally found (Figure 6a).

2.- Computer model for the interaction of PCI mutant forms V38G, V38A, V38L and V38I with CPA.

The mutant forms of PCI V38G,V38A,V38L and V38I were modelled over the wtPCI conformation (see above) in order to further assess the above discussed hypothesis on the binding affinities of the different PCI variants in residue 38. Their differences in K_i were then interpreted exclusively by the changes in interactions involving side chain atoms of residue 38 with the active site cavity of CPA and the hydrophobic effect of the substitution. After energy minimisation the

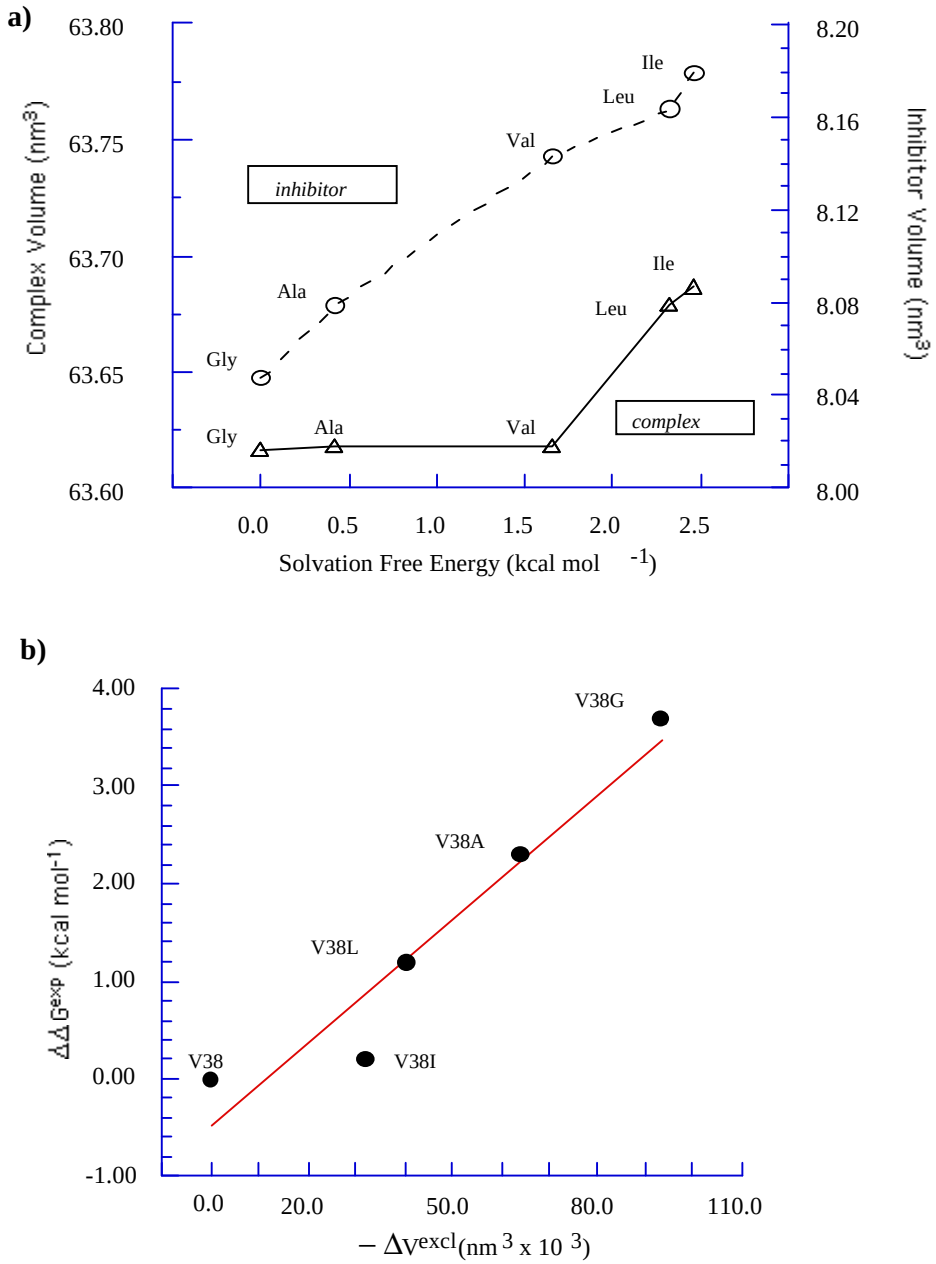


Fig. 7 *(a) Volumes of wtPCI and mutants, alone (right) and in complexes with CPA (left), versus the solvation free energies of mutated residue 38 of PCI, according to Eisenberg and Mc Lachlan (1986).*
(b) increase of ΔG_d^o of CPA-PCI complexes versus the difference of excluded volumes.

surface-triangulation program GEPOL [22] was used to calculate the volumes of wt and mutant PCIs and of their complexes with CPA. The excluded volume when forming the complex, V^{excl}, is calculated according to the expression:

$$V^{excl} = VCPA + VCPCI - VCPA - PCI$$

where V_{CPA} is the volume of CPA alone, V_{PCI} is the volume of PCI alone, and $V_{CPA\text{-}PCI}$ is the volume of the complex.

Computer graphic analysis of the energy minimised complexes shows that the electrostatic interactions and hydrogen bonds between both molecules are essentially maintained in the whole series. This is an obvious consequence of the model building employed. Subsequently, we compared the changes in volume with the variation in hydrophobicity, that is, with the tendency of the side chain to become buried, using the solvation free energy values of Eisenberg and McLachlan [29]. As shown in Figure 7a, there is a clear relationship between the volume increase of PCI alone and the corresponding increase in the solvation free energy of the different residues placed at position 38 of PCI. In contrast, the total volumes of the PCI-CPA complexes are only affected for mutant residues Ile or Leu, that is for those containing δ methyl groups. This also explains how the free energy of hydration affects the total variation on free energy exchange, mainly due to the exposure of the δ methyl groups to the solvent.

If our approach is correct there should be a linear relationship between the change in dissociation energy of the complexes Δ(ΔG) relative to the wtPCI-CPA complex and the increase in the excluded volume. We can calculate this change according to equation 6:

$$\Delta(\Delta G) = -k\left(V_m^{excl} - V_{wt}^{excl}\right) = -k\Delta V^{excl} \qquad (6)$$

being k = KTr [23] ; V_m^{excl}, the excluded volume when forming the complex CPA-mutant; V_{wt}^{excl}, the excluded volume when forming the complex wtPCI-CPA, and ΔV^{excl}, the difference in the excluded volume.

A linear relationship was found between Δ(ΔG) of the complexes calculated from the experimental data and the increase in the excluded volume for the different PCI variants (Figure 7b). Assuming that approximately r = 0.03 mol·Å^{-3} (molecular density of water alone) and T=300 K, we get k ≈ 0.02 (kcal·mol^{-1}K^{-1}Å^{-3}), which lies whiten the same order of magnitude than the one found for the slope from Figure 7b (m=0.04). All these results are in agreement with the hypothesis made to explain the experimental results, that any side chain aliphatic Cδ in position 38 is not buried in the active site cavity of CPA, probably being in contact with the water shell of the complex and not making any contribution to its stability.

Analysis of the generated structures of the different mutant inhibitor-CPA complexes in computer graphics helps to visualise and understand our results. The S1 subsite position in the enzyme, which is located in a narrow passage close to residue 38 of PCI, is followed by a wide pocket. Only chemical groups in positions β and γ of residue 38 would be in appropriate positions to be buried in the active centre of the CPA, while Cδ in mutants V38I and V38L are faced to the water shell. In the case of mutant V38L, one of the γ methyl groups of Val38 is absent. In wtPCI, this γ methyl interacts with the aromatic rings of two CPA residues: Tyr198 (at 3.79 Å) and Phe279 (at 3.81 Å). In the case of mutant V38I, in which both γ methyl groups are present, the K_i was similar to that of wtPCI. Figure 8 shows a stereo view of one model built mutant of PCI, Ile38, in the S1 subsite of CPA. It can be seen that the δ methyl group of the mutant is directed backwards to the PCI core, specifically to Trp28 of PCI, and has no significant contacts with CPA. Therefore, it cannot contribute to the stability of the complex.

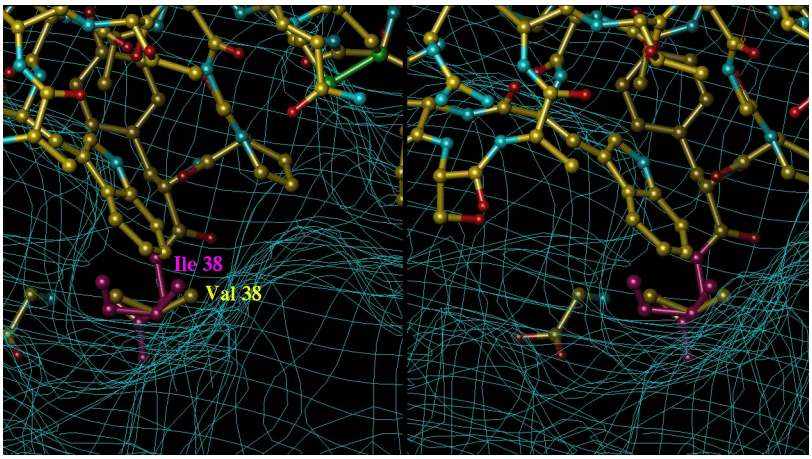


Fig. 8 *Computer graphic visualisation of the model-built mutant PCI V38I interacting with CPA in the S1 subsite. The stereo view depicts the spline Connolly surface (rolling sphere 1.4 Å) of CPA in blue.*

Conclusions

Despite its inaccuracy, the experimental results corroborated the predictions made by MD simulation for the C-tail mutants P36G and Y37F, that is, both mutants P36G and Y37F remained good inhibitors. It has to be kept in mind that the wide confidence interval found in the K_i determinations is quite normal in assays of inhibitory activities [7,15]. As a consequence, the K_i experimental values are not accurate enough to allow an exact quantification and comparison to those coming from molecular dynamics, which are also an approximation. However, the fact that the K_i (or ΔG_d^o) of the mutant P36G (designed to enhance C-tail mobility) is quite similar to that of the wild-type clearly indicates that proline 36 does not play such an important role for the inhibitory activity of PCI as was previously considered [11]. Although the present approach is still a qualitative one, in that a simplified molecular dynamics procedure is utilised, on many occasions this approach will be helpful to protein designers. At the present state of computer simulation methodology, it is generally assumed that it is quicker to create a site-directed mutant than to analyse it by MD. This is true when the replacement leads to a mutant polypeptide expressed as a soluble protein, as it was here, but when a mutation affects the folding or structural stability, obtaining the mutant polypeptide can be time-consuming or even impossible. This was the case for the G35P+P36G mutant, where we predicted an important variation on the functionality of the mutant. We found when trying to obtain the mutant that the net of hydrogen bonds found around residue 35 is also probably involved in the folding of PCI. Except for the case of mutations involved in the folding of a protein, the MD methodology in its present state may become not only competitive but an indispensable tool.

We have also shown that an MD simulation is not always needed. Thus, it is not an standard recipe for the theoretical study of mutations on proteins but a tool for analysis of its properties. As it is shown on our second approach for analysing the experimental results obtained in the case of mutants on residue 38 of PCI, an MD simulation would not have been helpful because the free energy exchange of the inhibition is related to the interaction of residue 38 with its environment, including the hydration energy of the molecules and the variation of contacts within the active site cavity of CPA. The most common method used for this kind of analysis, and also expensive on calculation, is the Free Energy Perturbation methodology [30]. Instead of this, we have taken advantage of the experimental knowledge of the protein used. We have tried to elude several long FEP calculations by a simple hypothesis to model the inhibitory mechanism of the mutant forms of PCI on residue 38. This approach has been successful with a simple refined model.

Acknowledgements. F. Canals acknowledges a post-doctoral fellowship from the Ministerio de Educación, Spain. C. Marino is a fellowship recipient from the Programa de Cooperación Científica con Iberoamerica. X. Daura is a fellowship recipient from the Programa de Química Fina (CICYT). This work has been supported by grants BIO94-0912-C02, BIO92-0458 and IN94-0347 from the CICYT (Ministerio de Educación, Spain) and by Fundación Francisca de Roviralta. The authors acknowledge CRAY computing facilities provided by CESCA.

References

1. Bowie J.V., Reidhaar-Olson J.F., Lim W.A., Sauer R.T., *Science* **1990**, *247*, 1306-1310
2. van Gunsteren W.F. and Berendsen H.J.C., *Angew. Chem. Int. Ed. Engl.* **1990**, *29*, 992-1023
3. van Gunsteren W.F. and Mark A.E. , *Eur. J. Biochem.* **1992**, *204*, 947-961.
4. Tapia O., Oliva, B., Nilsson O., Querol, E. and Avilés, F.X. , *J. Mol. Engineer* **1991**, *1*, 249-266.
5. Navia M.A., McKeever B.M., Springer J.P., Lin T.Y., Williams H.R., Fluder E.M., Dorn C.P., Hoogsteen K. *Proc. Nat. Acad. Sci. USA.* **1989,** *86*, 7-11.
6. Clore, G.M., Groneborn, A.M., Nilges, M. and Ryan, C.A. , *Biochemistry* **1987**, *26*, 8012-8023.
7. Rees, D.C. and Lipscomb, W.N. , *J. Mol. Biol.* **1982**, *160*, 475-498.
8. Hass, G.M., Ako, H., Grahn, D.G. and Neurath, H. , *Biochemistry* **1976**, *15* , 93-100.
9. Molina, M. A., Marino, C., Oliva, B., Avilés , F. X. and Querol, E. *J. Biol. Chem.* **1994,** *269*, 21467-21472 .
10. Molina M.A., Avilés F.X. and Querol E., *Gene* **1992**, *116*, 129-138.
11. Oliva, B., Wastlund, M., Cardenas, R., Querol, E., Avilés, F.X. and Tapia, O., *Biochem. Biophys. Res. Comm.* **1991**, *176*, 616-621.
12. Oliva, B., Wastlund, M., Nilsonn, O., C‡rdenas, R., Avilés, F.X., Querol, E. and Tapia, O., *Biochem. Biophys. Res. Comm.* **1991**, *176*, 627-632.
13. Marino-Buslje, C., Molina, N.A., Canals, F., Avilés, F.X. and Querol, E., *Appl. Microbiol. Biotechnol.* **1994**, *41*, 632-637.
14. Nakamaye, K.L. and Eckstein, F. *Nucl. Acids Res*. **1986**, *14*, 9679-9698.
15. Hass, J. M. and Ryan, C.A. *Carboxypeptidase Inhibitor from Potatoes*; Meth.Enzymol.: **1981**, *80* , pp 778-791.
16. Chang, J.Y., Canals, F., Schindler, P., Querol, E. and Avilés, F.X., *J. Biol, Chem.* **1994**, *269*, 22087-22094.
17. Henderson, P.J.F. *Biochem. J.* **1972**, *127*, 321-333.
18. van Gunsteren W.F. and Berendsen H.J.C. *Groningen Molecular Simulation (GROMOS) Library Manual.*; BIOMOS B.V.: Groningen., **1987**
19. van Gunsteren W.F. and Berendsen H.J.C. *Mol. Phys.* **1977**, *34*, 1311-1327.

20. Roussel, A. and Cambillau, C. *Silicon Graphics Directory* (Silicon Graphics Mountain View): California, USA, **1991.**
21. Åqvist, J., van Gunsteren W.F, Leijonmarck M., Tapia O.A., *J. Mol. Biol.* **1985**, *183*, 461-477.
22. Pascual-Ahuir, J. L., Silla, E., Tomasi, J. and Bonacorsi, R. *J. Comp. Chem.* **1987,** *8* , 778-787.
23. Richmond, T. J. *J. Mol. Biol*. **1984,** *178*, 63-89.
24. Dayhoff, M.O. In: *Atlas of Protein Sequence and Structure* ; Dayhoff, M.O., ed, National Biomedical Research Foundation, Washigton.: **1979,** vol.5, Suppl.3.
25. Bordo, D.,Argos, P. *J.Mol.Biol.* **1991**, *21*, 7721-729.
26. Todhunter J.A. *Methods in Enzymol.* **1979**, *63*, 391-393.
27. Fersht A.R. *Enzyme Structure and Mechanisms.* ; W.H. Freeman and Co.: New York. **1985**
28. Hass, G. M, and Hermodson, M. J. *Biochemistry* **1981**, *20*, 2256-2260.
29. Eisenberg, D. and McLachlan, A.D. *Nature* **1986**, *319*, 199-203.
30. Mark, A.E. and van Gunsteren, W.F. *J. Mol. Biol.* **1994**, *240*, 167-176.