ELSEVIER

Biophysical Chemistry 105 (2003) 29–43

Biophysical Chemistry

www.elsevier.com/locate/bpc

Homology modeling and calculation of the cobalt cluster charges of the *Encephazlitozoon cuniculi* methionine aminopeptidase, a potential target for drug design

François Bontems[a,*], Pascal le Floch[b], Francis Duffieux[a], Corinne Biderre[c,1], Pierre Peyret[c], Jean-Yves Lallemand[a]

[a]*Equipe ICSN-RMN, Ecole Polytechnique, F91128 Palaiseau France and Institut de Chimie des Substances naturelles, Av. de la Terrasse, Gif-sur-Yvette F91198, France*

[b]*Laboratoire Hétéroélements et Coordination UMR CNRS 7653, Département de Chimie, Ecole Polytechnique, Palaiseau F91128, France*

[c]*Institut Universitaire Technologique, Université d'Auvergne, Campus des Cézeaux, Aubière Cedex F63177, France*

Received 22 January 2003; received in revised form 7 March 2003; accepted 7 March 2003

Abstract

Fumagillin is a potent anti-angiogenic drug used in cancer treatments. It is also one of the few molecules active against the *Enterocytozoon* and *Encephalitozoon* parasites responsible for various clinical syndromes in HIV-infected or immunosuppressive treated patients. Its toxicity, however, makes desirable the design of more specific molecules. The fumagillin target, as anti-angiogenic agent, is the methionine aminopeptidase, an ubiquitous metallo-enzyme responsible for the removing of the N-terminal methionine in nascent proteins. By analogy, it has been proposed that this enzyme could also be the target in the parasites. As a first approach to verify this and to determine if it would be possible to design a more specific derivative, we have built a homology model of the *E. cuniculi* aminopeptidase. The charges of the two cobalt ions present in the active site and of the side-chains involved in their binding were computed using ab-initio methods. A preliminary comparison of the interactions of the fumagillin and of a related compound, the TNP-470, with both the human and the parasitic enzymes strongly support the hypothesis that the parasitic aminopeptidase is indeed the target of the fumagillin. It also suggests that the TNP-470 interact identically with both enzymes while there could be small differences in case of the fumagillin.

Keywords: AIDS; Fumagillin; Metallo-enzyme; Microsporidium; Parasite

*Corresponding author. Tel.: +33-1-69-33-48-55; fax: +33-1-69-33-30-10.
E-mail address: francois.bontems@polytechnique.fr (F. Bontems).
[1] Present address: Biogemma, Campus des Cézeaux, F63177, Aubière Cedex, France.

0301-4622/03/$ - see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0301-4622(03)00056-5

1. Introduction

The microsporidia are obligate intracellular parasites of mammals, including human [1]. Those parasites are responsible for various digestive and nervous clinical syndromes in HIV-infected or immunosupressive treated patients [2]. Microsporidiosis cases have been also detected in immunocompetent patients [3,4] and the estimate of the prevalence in the European population is approximately 8% [5]. Species involved in human diseases are mainly *Enterocytozoon bieneusi* and those of the *Encephalitozoon* genus (*E. cuniculi*, *E. Hellem*, *E. intestinalis*).

Only a few drugs were found to be effective against these microorganisms, the albendazol [6] the fumagillin [7] and several of its derivatives. Albendazol is a broad-spectrum anti-protozoal and anti-helminthic compound, which acts by inhibiting the microtubule polymerization. However, it is not very effective on *E. bieneusi* and on *Encephalitozoon* species and may lead to relapse [8] or incomplete response [9]. Fumagillin is an anti-amebiasis drug and an inhibitor of tumor-induced angiogenesis. Unfortunately, it is a highly toxic compound. Its synthetic analogue TNP-470 (O-[chloroacetylcarbamoyl]fumagillol) seems to be a more promising molecule, being less toxic and 50 times more active against tumors [10]. However, its potentialities as an anti-parasitic drug have not been really evaluated and the long-term consequences of a prolonged exposition are not clear, considering, in particular, its potent anti-angiogenic activity [10]. Moreover, while the mammalian molecular target of TNP-470 and fumagillin has been identified as the methionine aminopeptidase 2 (MetAP-2: EC 3.4.11.18) [11], the parasitic target is not known. Based on the recent identification of a gene product homologous to MetAP-2 in the *E. cuniculi* genome [12], we wondered if this protein could be the microsporidial target of TNP-470 and fumagillin. We were also interested in exploring the feasibility of designing new fumagillin derivatives, which could be more specific of the parasitic enzyme.

The computational approach will certainly not be sufficient to give an answer to these questions, but has proven to be very helpful to orient the experimental research [13,14]. A pre-requisite to approach these problems is to have a precise description of the protein structure and of its interactions with the drugs. The three-dimensional structures of the human aminopeptidase complexed with both the fumagillin and the TNP-470 have been reported [15], but the structure of the parasitic enzyme remains unknown. In addition, methionine aminopeptidases are metallo-enzymes. They possess two cobalt ions in their active site. The correct description of the electrostatic interactions between the enzymes and their inhibitors requires the calculation of the charges carried by the cobalt ions and by the residues involved in their chelation.

In this paper, we report the homology modeling of the *E. cuniculi* aminopeptidase three-dimensional structure and the calculation of the charges carried by the different elements involved in the cobalt cluster. This was used to obtain a preliminary comparison of the possible interactions between the drugs and both the human and the parasitic enzymes.

2. Material and methods

2.1. Soft- and hardware

The MODELLER [16], PROCHECK-NMR [17] and X-PLOR [18] softwares were implemented on either a Linux Personal Computer or a SUN Sparc20 workstation. The GAUSSIAN 98 [19] software runs on a DEC workstation. Graphical analyses of the structures were performed with MOLMOL [20] running on Linux or INDY SiliconGraphics workstations.

2.2. Model building

The open reading frame encoding the *E. cuniculi* MetAP-2 is located on the chromosome X (EMBL accession number: AL590449, [12]) from nucleotide 102630 to 103703 (gene number 10_0750).

All methionine aminopeptidase structures were downloaded form the RCBS (www.rcsb.org/pdb/) server. As explained in the Results section, three of them, namely the human enzyme 1B6A [15], the *Escherichia coli* homolog 1C24 [21] and the *Pyrococcus furiosus* protein 1XGS [22] were

retained for model building. They were aligned using the ALIGN3D routine of MODELLER. The alignment was further refined by visual inspection and by comparing the residue-by-residue Ramachandran plots calculated with PROCHECK-NMR. The *E. cuniculi* sequence was manually aligned with those of the three known structures. Two families of height models were then calculated using the standard MODELLER procedure. Their quality was verified by PROCHECK-NMR.

2.3. Charges calculation

Charge calculations were performed with the GAUSSIAN 98 suite of programs.

Geometry optimizations of fumagillin and TNP-470 were achieved at the Hartree-Fock (HF) level of theory using the 6-31G* basis set for all atoms (H, C, O, Cl). Vibrational frequencies of the stationary points were calculated at HF/6-31G* with numerical second derivatives of the energy with respect to the coordinates. The optimized structures were found to be located at minimum on their potential-energy surfaces.

The charges of the cobalt cluster were determined using the coordinates of both the 1BOA and 1B6A PDB files [15]. In each case, the two cobalt ions, the side-chains of the Asp251, Asp262, His331 (truncated at the C^{α} position) Glu364 and Glu459 (truncated at the C^{β} position), and a water molecule (HOH698 in the case of 1B6A, HOH662 in the case of 1BOA) were introduced in the model. A single point calculation was carried at the B3LYP level of theory, which consists of a hybrid Becke+Hartree-Fock exchange and Lee-Yang-Par correlation functional with non-local corrections [23]. A relativistic effective core potential (ECP) of Way and Wadt with a (421/211/41) split-valence basis set was used for the cobalt [24] while the 6-31G* basis set was used for the other (H, N, C, O) atoms. The bonding situations of the experimental structures were analyzed using the natural bond orbital (NBO) method developed by Weinhold [25].

2.4. Inhibitor force field derivation

A force field was derived for the fumagillin and the TNP-470. A complete list of bond, angle, dihedral and improper terms was generated for each molecule by X-PLOR. The reference bond length and angle, dihedral and improper values were derived from the gaussian optimized geometries. A set of force strength constant was first derived by using the 'parameter learn' and 'parameter reduce' procedures of X-PLOR. This set was completed by using values found for similar terms in the CHARMM22 parameter file.

2.5. Docking and interaction analysis

Two of the modeled structures were conserved to analyze the interactions. They were superimposed with either the 1B6A or the 1BOA structures and the coordinates of the cobalt ions and of the corresponding inhibitors were transferred from the PDB to the model files. All protons were then built using the X-PLOR 'hbuild' command. The addition of the protons in both the initial crystallographic and the modeled structures results in many bad contacts, which were suppressed by minimization. A strong harmonic potential (1000 kcal mol^{-1} $Å^{-1}$) was applied to avoid too large deformation of the structures in the first step of the process. This restraint was gradually released. At the end, a last minimization was performed without restraint on protein atoms. However, a small value (50 kcal mol^{-1} $Å^{-1}$) was conserved to maintain the positions of the ions, of the residues bound to them, of the inhibitor atoms, and of the water molecule.

The electrostatic potential was calculated at the protein surface using the CalcProt command of the MOLMOL software. The solvent and protein dielectric constant were set to 80 and 2, respectively. A salt concentration of 0.3 mol l^{-1} was used. The AtomCharge file was modified in order to take account of all atom charges, including those calculated for the cobalt cluster.

3. Result and discussion

3.1. Prediction of the E. cuniculi aminopeptidase structure

3.1.1. Presentation of the existing structures

The methione aminopeptidases form a class of ubiquitous enzymes responsible for the removing

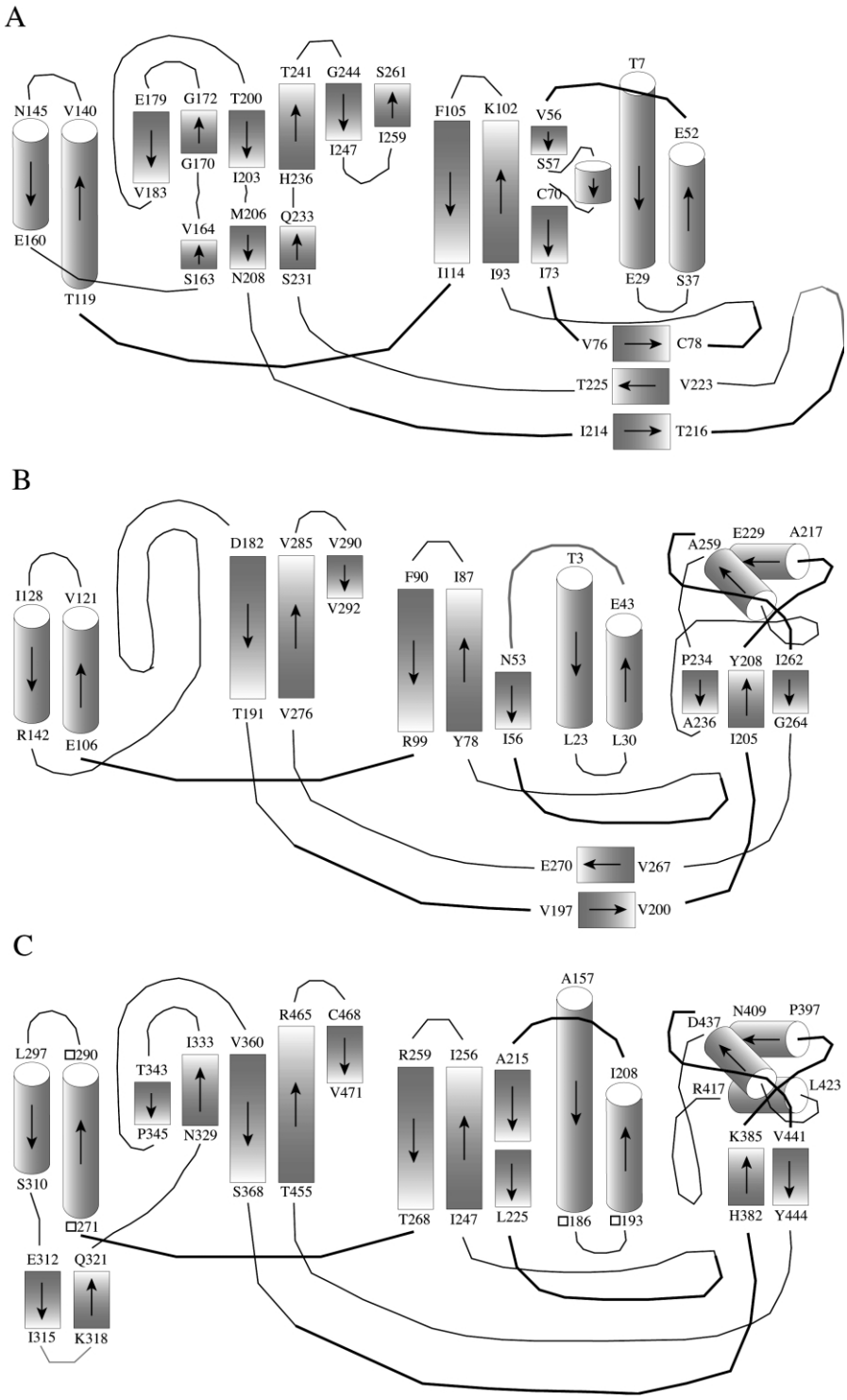
A
N145
V140
E179
G172
T200
T241
G244
S261
F105
K102
V56
T7
E52
E160
V183
G170
I203
H236
I247
I259
S57
C70
V164
M206
Q233
T119
S163
N208
S231
I114
I93
I73
E29
S37
V76
C78
T225
V223
I214
T216
B
I128
V121
D182
V285
V290
F90
I87
T3
E43
A259
E229
A217
R142
E106
T191
V276
V292
R99
Y78
N53
I56
L23
L30
P234
Y208
I262
A236
I205
G264
E270
V267
V197
V200
C
L297
□290
T343
I333
V360
R465
C468
R259
I256
A215
A157
I208
D437
N409
P397
S310
□271
P345
N329
S368
T455
V471
T268
I247
L225
□186
□193
R417
L423
K385
V441
H382
Y444
E312
Q321
I315
K318

Fig. 1.

of the *N*-terminal methionine from the nascent proteins. The analysis of their sequences suggests that they all are evolutionary related. This is confirmed by the comparison of the three-dimensional structures of the *E. coli*, human and *P. furiosus* enzymes. As shown in Fig. 1 and Fig. 3, they possess the same 'pita-bread' fold [26] formed of two repetitions of an $\alpha+\beta$ motif (two helices, at least three β-strands). In addition, a small triple-strand β-sheet made from two hairpin loops coming from the two halves of the pita-bread fold is always present. The examination of the Fig. 1, drawn from the indications provided by the PDB file headers, may suggest existence of significant variations of this core-motif (for example: in the number of strands involved in the β-sheets). However, this is only due to differences in the secondary structure element declaration, the direct observation of the three structures indicating a strict conservation of the topology (Fig. 3).

The presence of several variable elements beside this conserved core has led to a classification of these enzymes in two main groups. The major difference between class I (formed by the bacterial and several eukaryotic enzymes) and class II (archebacterial and other eukaryotic enzymes) is the presence of a 60 amino acid α/β subdomain inserted within the triple-strand β-sheet in the latter. In addition, eukaryotic class II enzymes possess a small hairpin inserted in the second half of the pita-bread fold (between the second helix and the first β-strand). Finally, they were also characterized by the presence of an additional *N*-terminal domain.

The active site is mainly composed of residues belonging to the two β-sheets of the pita-bread fold and to the triple-strand extension. It is located in the crevasse between the fold and the supplementary α/β domain in the class II enzymes. Five residues (Asp251, Asp262, His331, Glu364 and Glu459, using the human enzyme numbering) are involved in the chelation of two metal ions (two cobalt ions in the crystal), forming a binuclear cluster. Two additional histidines (His231 and His339) are thought to participate to the reaction. Finally, several side-chains are involved in the substrate recognition, Phe219, His331, Ile338, His339 and Tyr444 forming the putative methionine-binding pocket, for example [27].

3.1.2. Choice of the reference structures

The *E. cuniculi* aminopeptidase is a group II eukaryotic enzyme, and thus we could have modeled its structure using that of the human aminopeptidase alone. However, we preferred to use a more conservative strategy and to keep more than one template. Indeed, we considered that a model is not an experimental structure and is bound to contain incorrect regions, even when the sequences are conserved. In these conditions we preferred to work with an under-determined model than to attempt to draw conclusions from an over-determined one. We chose the MODELLER software for the same reason, as it can take into account the structural and sequence variability of the templates and generate a family of models instead of a single one, allowing the direct observation of the well and the less-well predicted regions.

Several structures were available for each of the three structurally studied aminopeptidases: nine for the *E. coli* enzyme (PDB entries 1C21, 1C22, 1C23, 1C24, 1C27, 1MAT, 2MAT, 3MAT and 4MAT), four for the human (1B6A, 1BOA, 1BN5 and 1B59) and three for the *P. furiosus* (1XGM, 1XGO and 1XGS). The cobalt ions are absent or localized outside the active site of the 1MAT,

Fig. 1. Comparison of three methionine aminopeptidase topologies. (a) *E. coli* (PDB entry 1C24); (b) *P. furiosus* (1XGS) and (c) *H. sapiens* (1C24). The diagrams have been drawn using the indications reported in the PDB headers. These indications suggest that some β-strands are not conserved (the V76-C78, V223-T225 and T214-T216 strands of the *E. coli* enzyme, for example). This, however, is refuted by the visual inspection of the three structures. The core domain is, indeed, very similar in the three proteins. It is mainly composed of a pita-bread fold, (best represented in A by the T7-T241 region) and of a triple-stranded β-sheet (N208-S231 and I73-I93 region, also in A) The main difference between the *E. coli* (class I) and the two other (class II) structures is the insertion of an α/β (I205-G264 in B) at the apex of one of the loops forming the triple-strand region. Among the class II enzymes, it is possible to further discriminate the eukariotic and archebacterial structures, the former possessing a small β-hairpin (E312-Q321 in C) inserted between the second helix and the first strand of the second half of the pita-bread fold.

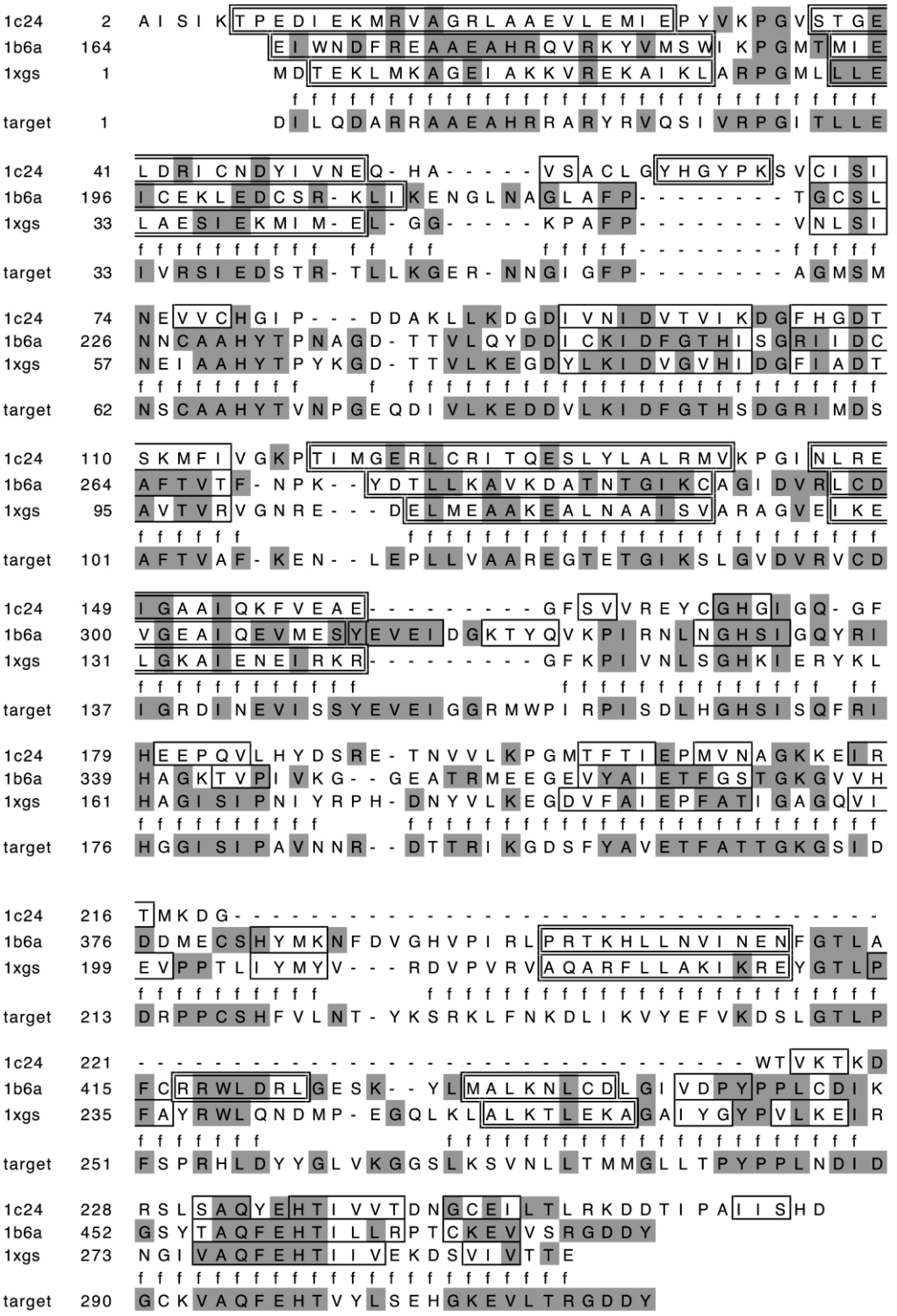
1c24 2 AISIKTPEDIEKMRVAGRLAAEVLEMIEPYVKPGVSTGE
1b6a 164 EIWNDFREAAEAHRQVRKYVMSWIKPGMTMIE
1xgs 1 MDTEKLMKAGEIAKKVREKAIKLARPGMLLLE
target 1 DILQDARRAAEAHRRARYRVQSIVRPGITLLE
1c24 41 LDRICNDYIVNEQ-HA-----VSACLGYHGYPKSVCISI
1b6a 196 ICEKLEDCSR-KLIKENGLNAGLAFP--------TGCSL
1xgs 33 LAESIEKMIM-EL-GG-----KPAFP--------VNLSI
target 33 IVRSIEDSTR-TLLKGER-NNGIGFP--------AGMSM
1c24 74 NEVVCHGIP---DDAKLLKDGDIVNIDVTVIKDGFHGDT
1b6a 226 NNCAAHYTPNAGD-TTVLQYDDICKIDFGTHISGRIIDC
1xgs 57 NEIAAHYTPYKGD-TTVLKEGDYLKIDVGVHIDGFIADT
target 62 NSCAAHYTVNPGEQDIVLKEDDVLKIDFGTHSDGRIMDS
1c24 110 SKMFIVGKPTIMGERLCRITQESLYLALRMVKPGINLRE
1b6a 264 AFTVTF-NPK--YDTLLKAVKDATNTGIKCAGIDVRLCD
1xgs 95 AVTVRVGNRE---DELMEAAKEALNAAISVARAGVEIKE
target 101 AFTVAF-KEN--LEPLLVAARFGTETGIKSLGVDVRVCD
1c24 149 IGAAIQKFVEAE----------GFSVVREYCGHGIGQ-GF
1b6a 300 VGEAIQEVMESYEVEIDGKTYQVKPIRNLNGHSIGQYRI
1xgs 131 LGKAIENEIRKR----------GFKPIVNLSGHKIERYKL
target 137 IGRDINEVISSYEVEIGGRMWPIRPISDLHGHSISQFRI
1c24 179 HEEPQVLHYDSRE-TNVVLKPGMTFTIEPMVNAGKKEIR
1b6a 339 HAGKTVPIVKG--GEATRMEEGEVYAIETFGSTGKGVVH
1xgs 161 HAGISIPNIYRPH-DNYVLKEGDVFAIEPFATIGAGQVI
target 176 HGGISIPAVNNR--DTTRIKGDSFYAVETFATTGKGSID
1c24 216 TMKDG----------
1b6a 376 DDMECSHYMKNFDVGHVPIRLPRTKHLLNVINENFGTLA
1xgs 199 EVPPTLIYMYV---RDVPVRVAQARFLLAKIKREYGTLP
target 213 DRPPCSHFVLNT-YKSRKLFNKDLIKVYEFVKDSLGTLP
1c24 221 ----------WTVKTKD
1b6a 415 FCRRWLDRLGESK--YLMALKNLCDLGIVDPYPPLCDIK
1xgs 235 FAYRWLQNDMP-EGQLKLALKTLEKAGAIYGYPVLKEIR
target 251 FSPRHLDYYGLVKGGSLKSVNLLTMMGLLTPYPPLNDID
1c24 228 RSLSAQYEHTIVVTDNGCEILTLRKDDTIPAIISHD
1b6a 452 GSYTAQFEHTILLRPTCKEVVSRGDDY
1xgs 273 NGIVAQFEHTIIVEKDSVIVTTE
target 290 GCKVAQFEHTVYLSEHGKEVLTRGDDY

Fig. 2.

3MAT 4MAT and 1XGO structures. They were thus discarded. A visual analysis of the others showed that all structures corresponding to the same protein are nearly identical. The root mean square deviation (rmsd) calculated on the backbone atoms are 0.2, 0.2 and 0.4 Å for the *E. coli*, human and *P. furiosus* proteins, respectively. Things become more complicated when considering the side-chains, as several of them are disordered (21 side-chains are missing in, at least, the C2 files, 12 being absent in all of them). However, the rmsd calculated on all conserved heavy atoms remain low: 0.3, 0.5 and 0.7 Å.

Considering this, we decided to retain the best structure for each protein, namely the 1C24 (resolution: 1.70 Å), the 1B6A (1.60 Å) and the 1XGS (1.75 Å) for the *E. coli*, human and *P. furiosus* aminopeptidase, respectively.

3.1.3. Structural and sequence alignment

We first aligned the two class II (1B6A and 1XGS) structures using the ALIGN3D routine of the MODELLER software. They appear very similar: the positions of 282 residues are conserved, with a global rmsd calculated on the backbone atoms equal to 1.3 Å. The main differences concern the replacement of the Ile208-Ala215 loop in the eukaryotic structure by the longer Glu43-Asn53 loop in the archebacterial protein and the insertion of the Glu312-Gln321 hairpin between the second helix and the first strand of the second half of the pita-bread fold in the eukaryotic protein. The small α/β inserted domain also appears more variable than the remaining of the structure.

Similarly, we also used MODELLER to superimpose the three 1B6A, 1XGS and 1C24 structures. This structural alignment was refined both by visual inspection and by comparing the amino-acid-by-amino-acid Ramachandran plots calculated with PROCHECK-NMR for each structure. These plots were in particular very useful to precisely determine the end of the conserved fragments and to verify the alignment of the amino acids having an α_L conformation. The conserved segments concern less residues (217 instead of 282), but this is mainly due to the absence of the α/β subdomain in the prokaryotic structure. The only other important difference is the insertion in the *E. coli* structure of a small α-helical region within the first β-strand of the first $\alpha+\beta$ motif. But, this insertion has nearly no influence on the geometry of the remaining of the considered β-strand.

The *E. cuniculi* sequence was manually aligned on the three others (Fig. 2). There was nearly no insertion and deletion outside those previously discussed. The identity scores calculated between the *E. cuniculi* and the *E. coli*, *P. furiosus* and human sequence is 20, 38 and 52%, respectively. This score is even higher (58%), if we count the positions conserved between the *E. cuniculi* sequence and at least one of the three others. The worst score (30%) is obtained within the α/β subdomain.

3.1.4. Comparison of the two families of predicted structures

Two series of eight structures were calculated using the two class II enzymes as template for the first and the three available structures for the second. They are presented on Fig. 3.

In both cases, the eight structures are very similar as shown by the rmsd calculated on the backbone and heavy atoms. The average rmsd values between all pairs of structures are 0.42 ± 0.09 (backbone) and 1.73 ± 0.08 (heavy atoms) for the first series and 0.33 ± 0.06 (backbone) and 1.64 ± 0.06 (heavy atoms) for the second. The largest variability is observed in the α/β subdomain in both cases, but several short segments are also poorly constrained, such as the Gly73-Ile77 or the Asn186-Asp188 regions.

The two families are very similar. The rmsd, calculated as previously, are 0.39 ± 0.09 (backbone) and 1.70 ± 0.10 (heavy atoms). However,

Fig. 2. Alignment of the four *H. sapiens* (1b6a), *E. coli* (1c24), *P. furiosus* (1xgs) and *E. cuniculi* (target) aminopeptidase sequences. The alignment of the 1c24, 1b6a and 1xgs sequences was derived from the structure superimposition, as explained in Material and methods. The β-sheet and α-helix, as defined in the PDB file headers, are delineated by simple and double boxes, respectively. The structurally conserved regions are indicated by an 'f' under the sequences. The target sequence was manually aligned. The amino-acids shared by the target and, at least, one of the three other sequences are in gray.

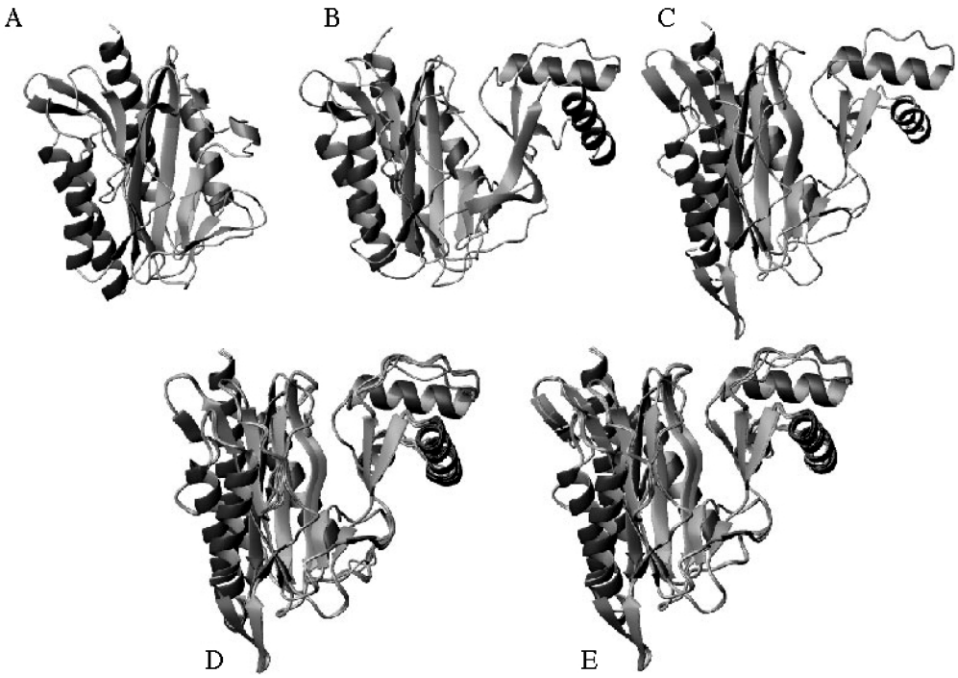

Fig. 3. Comparison of the modelized and crystallographic structures. The three template structures are represented in (A) (1C24, *E. coli*); (B) (1XGS, *P. furiosus*) and (C) (1XGS, *H. sapiens*). The two families of model are in D (models calculated by using the 1B6A and 1XGS structures) and in E (models calculated by using the three X-ray structures). The secondary structure elements have been calculated with MOLMOL. Thus, they do not exactly correspond to those presented in the Fig. 1.

several loops, such as the Gly47-Asn50, Glu74-Ile77 and Asn185-Asp188 fragments have different conformations in the two families. This indicates that the introduction of the class I enzyme in the calculation has an influence but, this influence is moderate and confined to small regions of the model.

3.1.5. Charge calculation

The investigation of the interactions between the different forms of the enzyme and their inhibitors requires a precise representation of the non-bonded, and in particular of the electrostatic, energy term. In case of the inhibitors (fumagillin and TNP-470), we used the natural charges determined after a full optimization of the geometry (Fig. 4). In case of the protein, the simplest solution would have been to use the standard CHARMM22 charges for the residues and to add two 2^+ charges for the two cobalt ions. However, the presence of the ions is bound to influence their surrounding, and vice-versa. We thus decided to calculate the charges carried by the two cobalts and by the residues involved in their chelation (including a water molecule). We used two sets of coordinates derived from the 1BOA (aminopeptidase complexed with fumagillin) and 1B6A (aminopeptidase complexed with TNP-470). The side-chains of the Asp251, Asp262 and His331 were truncated at the C^{β} position, the C^{α} being replaced by a proton, whereas those of the Glu354 and Glu459 were truncated at the C^{γ} positions. The calculations were performed for both clusters in the absence and in the presence of the cobalt ions. The results are reported in Fig. 5.

The charges of the side-chain atoms are very similar in the two clusters. In the absence of the cobalt ions, the difference is smaller than 0.01 for 38 of the 43 atoms, the largest deviation concerning the Asp251 $O^{\delta 2}$ (0.0185), the His331 $H^{\delta 1}$

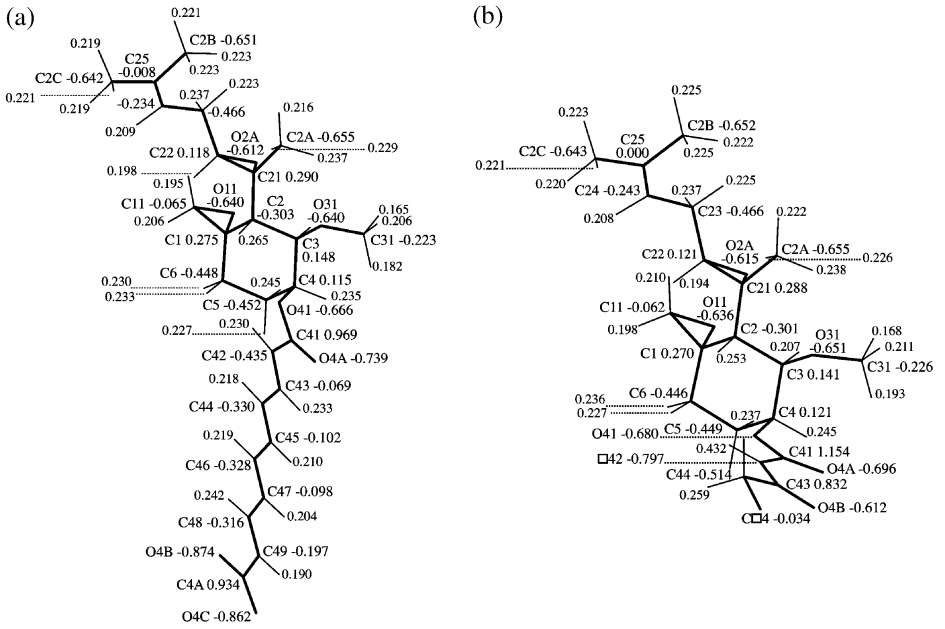

Fig. 4. Charges of the fumagillin (a) and TNP-470 (b) inhibitors. The nomenclature and numbering of all heavy atoms are indicated, they have been omitted for the hydrogens.

(0.0155), the Asp262 $O^{\delta 1}$ and $O^{\delta 2}$ (0.0130 both) and the water H^2 (0.0125). In the presence of the cobalt ions, the difference is smaller than 0.01 for 33 atoms and the largest deviations mainly concern the glutamate carbonyl functions (Glu354 $O^{\varepsilon 1}$: 0.0315; Glu459 $O^{\varepsilon 1}$: 0.0455; Glu459 $O^{\varepsilon 2}$: 0.0345). The results obtained for the cobalt ions themselves are more contrasted. The charges deviate significantly from 2^+ and are very different for the two ions (approximately 0.7 for one and 1.6 for the other). In addition, they were swapped between the two clusters. The first cobalt has a charge equal to 0.658 in the first cluster and equal to 1.545 in the second, while the second cobalt has a charge equal to 1.646 in the first cluster and equal to 0.700 in the second. The comparison of the cluster structures shows a very nice superimposition of the amino-acid side-chains (rmsd: 0.1 Å), but also indicates that the ions occupy slightly different positions. They move by approximately 0.3 Å . The differences between the two clusters thus seem due to very small displacements of the ions. It is likely that they will be averaged by the thermal movements at room temperature. Therefore, we have decided to use the same charge set for the two clusters, obtained by averaging the two calculated values for each atom (Fig. 6).

The comparison of the mean charge sets obtained in the presence and in the absence of the cobalt ions indicates that the ion main effect is to decrease, in algebraic value, the charge of the atom directly bound to it and to increase the charge of all the others. The negative effects are generally small (from -0.014 for the Asp251 $O^{\delta 2}$ to -0.044 for the Asp262 $O^{\delta 2}$), with the exception of the His331 $N^{\varepsilon 2}$ (-0.235). The positive effects are larger, the greatest value being that of the Asp262 C^{γ} (0.119). We also compared the charge set calculated in the absence of the cobalt ions with that of the CHARMM22 force field. We found a good agreement when looking at the charges carried by the functional groups (carbonyl and imidazol). However, this is no more true when the aliphatic protons and carbons are considered. The mean values of the charges carried by the aspartate C^{β} and H^{β} are -0.758 and 0.173 in our calculations, to be compared to -0.280 and 0.009 in CHARMM22. We thought that the truncation at

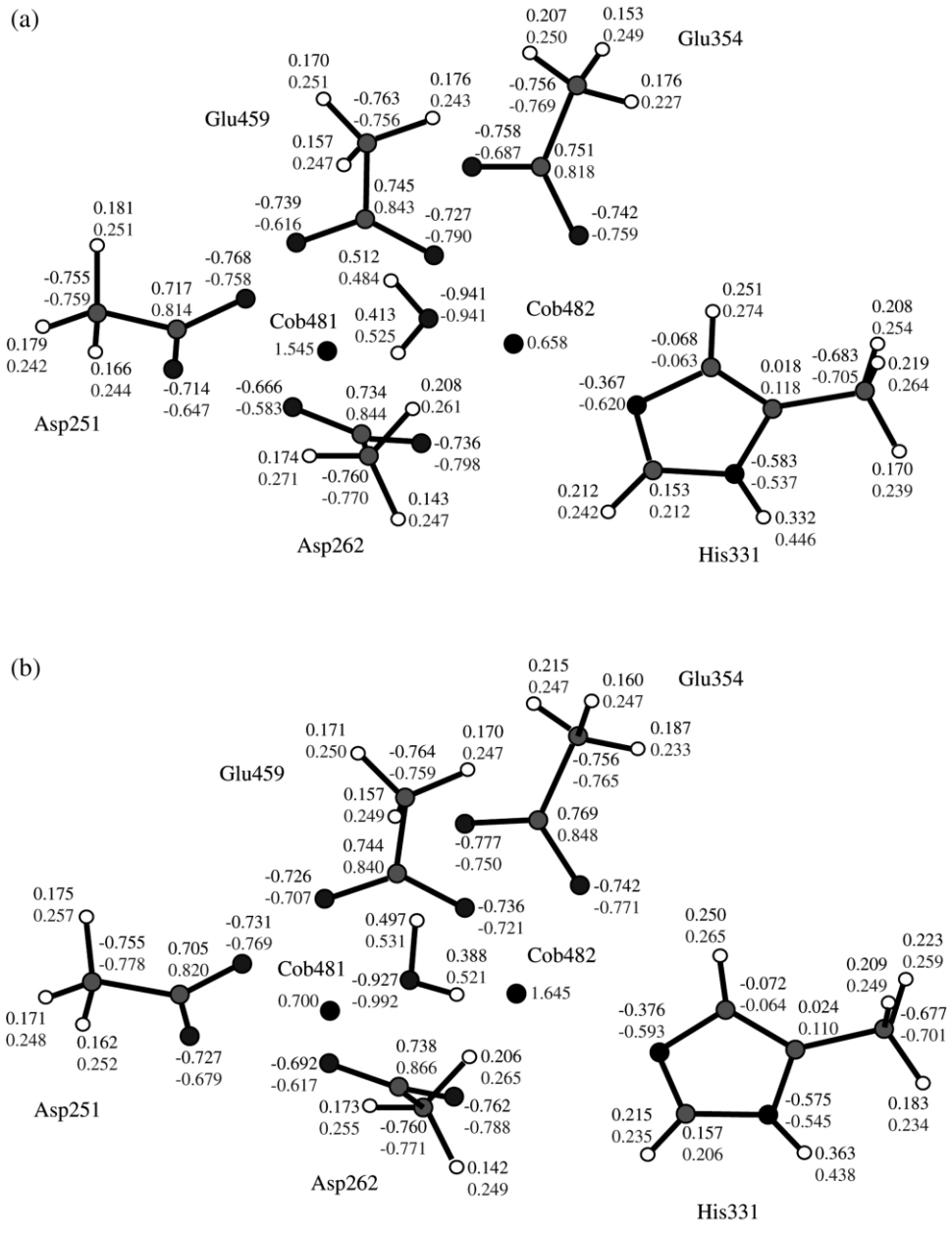

Fig. 5. Charges of the atoms involved in the cobalt cluster. Two sets of coordinates were used. The first (a) is issued of the 1B6A PDB file (aminopeptidase complexed with fumagillin), the second (b) of the 1BOA file (aminopeptidase complexed with TNP-470). The hydrogen positions have been added using the hbuilt routine of the X-PLOR software. The charges have been calculated in the absence (black, upper value) and in the presence (gray, lower value) of the cobalt ion.

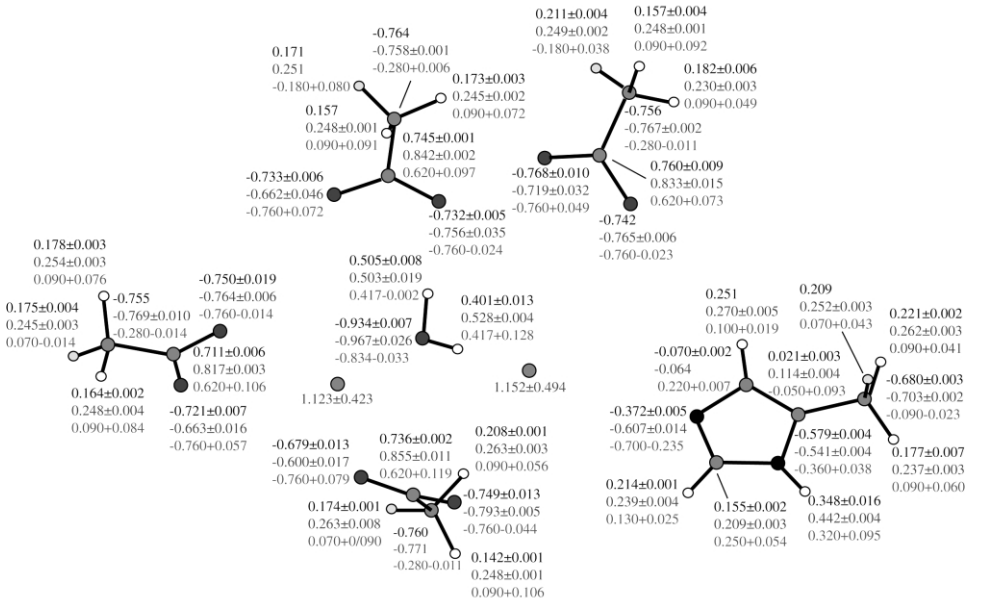

Fig. 6. Charges of the cobalt cluster atoms used in the energy calculation. For each atom, a correction corresponding to the difference between the mean values calculated in the presence and in the absence of the cobalt ions was added to the CHARMM charge. The upper and middle values are the mean values (± the variation) of the charge calculated in the absence and presence of the ions, respectively. The CHARMM22 value and the applied correction are the lower values.

the C^{β} or C^{γ} position was likely responsible of this discrepancy. However, a calculation carried out on a fully optimized model of aspartate residue (H_3N–CO–Asp–NH_3) ruled out this hypothesis. The charges of the H^{β} (0.225) atoms are very similar to those obtained for the cluster. That carried by the C^{β} (−0.570) is smaller than in the cluster, but remain higher than in CHARMM22. The simple replacement of the CHARMM22 charges by those obtained in our calculation would, thus, introduce a discontinuity at the truncation position. To avoid this, we decided to use the difference between the charge sets calculated with and without the cobalt ions as a perturbation and to add it to the CHARMM22 charge set (Fig. 6).

3.1.6. Inhibitor docking and interactions analysis

As previously discussed, the different models we computed have very similar backbone structures and the main differences are located in regions far from the active site. However, in order to analyze the interactions between the models and the inhibitors, the variability of the side-chain and, in particular, of those involved in the inhibitor recognition has to be considered. The analysis of the interactions between the fumagillin and the human aminopeptidase [21] reveals that the direct contacts between the protein and the inhibitor essentially involved nine side-chains. The fumagillin is covalently bound to the His231. The hydrophobic substituent of the C4 atom interacts with the Phe219, His331, Ile338, His339 and Tyr444. The C6 side-chain contacts the Leu328 and Leu447 through hydrophobic interactions, and the Asp376 through a hydrogen bond. All these residues are conserved in the *E. cuniculi* enzyme. The side-chain orientations vary according to the considered model. All orientations, but that of the Ile338, are conserved in the sixth model. Four side-chains possess different orientations in the first (Ile338, Leu328, Tyr444, Asp376), third (Ile338, His339, Tyr444, Asp376) and eighth (Phe219, Leu328, His339, Tyr444) models. We thus decided to dock the fumagillin and the TNP470 inhibitors in both the first and sixth models.

Table 1
Comparison of the van der Waals and electrostatic energy of the complexes before and after minimization All energy are in kcal/mol. All rmsd have been calculated on the heavy atoms, using directly the coordinates of the complexes taken before and after the minimization. They thus integrate both the deformation and the displacement of the considered elements. The structure quality has been evaluated, after minimisation, using PROCHECK. The global quality factor (G fac.), pecentage of the residues in the strictly allowed Ramachandran erea (Rama), the omega diedral angle standard deviation (ω std) and the number of bad contacts (Bad ct) are reported

Protein	Ligand	Before minimization			After minimization			RMSD			Procheck parameters			
		VdW Total	VdW Prot/lig	Eiec Prot/iig	VdW Total	VdW Prot/lig	Eiec Prot/lig	Protein	Ligand	Ions	G fac.	Rama (%)	(ω std)	Bad ct
1B6A	fumaqillin	23 714	24 539	−49	−1605	−43	−45	0.29	0.57	0.41	−0.1	91	6.8	0
1BOA	TNP-470	19 825	21 014	2	−1627	−43	−12	0.29	0.56	0.41	−0.1	91	6.8	0
Model 1	Fumaqillin	10 877	4531	103	−1232	−40	29	0.53	0.79	0.38	−0.4	90	7.9	0
Model 6	Fumagiilin	23 718	13 085	132	−1249	−46	26	0.48	0.69	0.37	−0.3	87	7.4	0
Model 1	TNP-470	10 385	4044	−15	−1233	−39	−25	0.53	0.66	0.41	−0.4	90	7.9	0
Model 6	TNP-470	11 781	1091	−39	−1249	−44	−42	0.48	0.56	0.40	−0.3	87	7.4	0

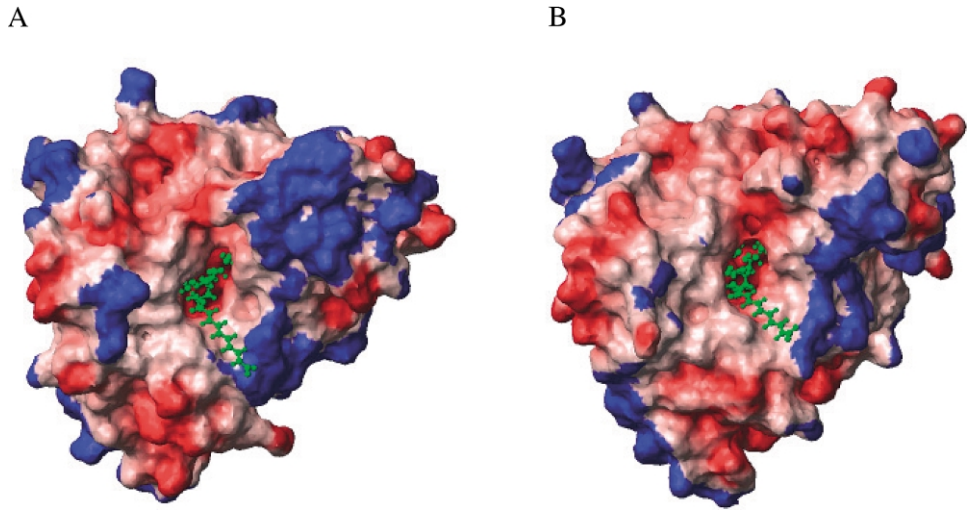

Fig. 7. Representation of the electrostatic field calculated at the surface of the crystallographic 1B6A (A) or of the modelized; (B) structures. The computational and display parameters were the same in both cases. The red and blue colors correspond to positive and negative potential, respectively. The structure of the fumagillin has been represented in green. The cyclic part of the inhibitor is embedded in a deep positive pocket in both cases, while the tail make contacts with the mouth of the pocket, in a region where the potential seems more variable.

The docking was simply realized by coordinate transfer. The conserved regions of the models were superimposed to those of either the 1BOA (with fumagillin) or 1B6A (with TNP-470) structures. The coordinates of the inhibitor, cobalt ions and water molecule were extracted from the PDB files and merged with those of the models. The six complexes (1BOA, 1B6A, model 1 with fumagillin, model 1 with TNP, model 6 with fumagillin and model 6 with TNP) were submitted to the same minimization protocol and the quality of the structures was verified using PROCHECK (Table 1).

All non-minimized complexes have a very high energy, due to a large van der Waals contribution (Table 1). This energy become negative in all cases after the minimization, the values being, however, better in case of the two crystallographic structures. This is not unexpected, considering that the side-chain packing in the models is not likely to be optimal. For the same reason, the deformation of the structure during the minimization is larger for the models (0.51 Å) than for the X-ray structures (0.21 Å), but both values are low. A large part of the initial van der Waals energy is due to bad contacts between the proteins and the inhibitors, the values being, surprisingly, worst for the crystallographic complexes than for the models. The values are very comparable at the end of the minimization with a very small advantage for the complexes involving the sixth model (-42 kcal/mol^{-1} instead of -36 for the first model and -40 for the X-ray structures).

The similarities of the structures after the minimization and of the van der Waals interaction energies indicate that the contacts made by both inhibitors are nearly identical in the crystallographic and in the modeled complexes. This is, indeed, confirmed by the visual inspection of the structures. However, the comparison of the electrostatic energies indicates that there is a small difference between the model and the X-ray structures. The electrostatic interaction between the crystallographic structure and the two inhibitors is always negative (-45 kcal mol^{-1} for the fumagillin, -12 for the TNP), This is also the case for the interaction of the TNP with both the first and the sixth model (-25 and -42 kcal mol^{-1}, respectively). However, this energy is positive in case of the fumagillin (29 and 26 kcal mol^{-1}). The

inspection of the electrostatic potential map (Fig. 7) allows a rationalization of this observation. The cyclic region of the fumagillin and of the TNP-470 interacts with a deep pocket of the enzyme, which strong positive potential seems identical for both the human and the parasitic enzyme. But, the fumagillin possesses, in addition, a long chain interacting with the mouth of the cavity, and the potential of this region seems more variable.

In conclusion, the model of the *E. cuniculli* aminopeptidase suggests that its active site is identical to the human one and that all residues directly involved in the interaction with either the fumagillin or the TNP-470 are structurally conserved. This strongly supports the hypothesis that the *E. cuniculli* class II aminopeptidase is the target of these two anti-parasitic drugs. The analysis of the electrostatic potential maps also suggests that the human and the parasitic enzymes interact in a similar manner with the TNP-470, but that there may be a slight difference in case of the fumagillin. Accordingly, the synthesis of a more selective inhibitor would require interactions outside the methionine-binding pocket. In addition, we think that the charge set we derived will be useful for further analysis of the *E. cuniculli* Met-aminopeptidase, but also, considering the conservation of the residues involved in the cobalt liganding, of the other members of this enzyme family.

All files (complex coordinates, inhibitor coordinates, topologies and parameters, ion cluster charges) are available on request to the corresponding author: (francois.bontems@polytechnique.fr).

References

[1] E.U. Canning, J. Lom, The Microsporidia of Vertebrates, Academic Press Inc, New York, NY, 1986.
[2] D.P. Kotler, J.M. Orenstein, Clinical syndromes associated with microsporidiosis, Adv. Parasitol. 40 (1998) 1–49.
[3] L. Raynaud, F. Delbac, V. Broussolle, M. Rabodonirina, V. Girault, M Wallon, et al., Identification of *E. intestinalis* in travelers with chronic diarrhea by specific PCR amplification, J. Clin. Microbiol. 36 (1998) 37–40.
[4] L. Cotte, M. Rabodonirina, F. Chapuis, F. Bailly, F. Bissuel, C. Raynal, et al., Waterborne outbreak of intestinal microsporidiosis in persons with and without human immunodeficiency virus infection, J. Infect. Dis. 180 (1999) 2003–2008.
[5] T. van Gool, J.C. Vetter, B. Weinmayr, A. Van Dam, F. Derouin, J. Dankert, High seroprevalence of Encephalitozoon species in immunocompetent subjects, J. Infect. Dis. 175 (1997) 1020–1024.
[6] E.D. Aarons, W.S. Woodrow, E.U. Hollister, E.U. Canning, N. Francis, B.G. Gazzard, Reversible renal failure caused by a microsporidian infection, AIDS 8 (1994) 1119–1121.
[7] J.A. Shadduck, Effect of fumagillin on in vitro multiplication of *E. cuniculi*, J. Protozool. 27 (1980) 202–208.
[8] R. Weber, B. Sauer, M.A. Spycher, P. Deplazes, R. Keller, R. Ammann, et al., Detection of *Septata intestinalis* in stool specimens and coprodiagnostic monitoring of successful treatment with albendazole, Clin. Infect. Dis. 19 (1994) 342–345.
[9] R. Weber, P. Deplazes, M. Flepp, A. Mathis, R. Baumann, B. Sauer, et al., Cerebral microsporidiosis due to *E. cuniculi* in a patient with human immunodeficiency virus infection, N Engl. J. Med. 336 (1997) 474–478.
[10] E.A. Kruger, D.W. Figg, TNP-470: an angiogenesis inhibitor in clinical development for cancer, Exp. Opin. Invest. Drugs 9 (2000) 1383–1396.
[11] N. Sin, L. Meng, M.Q. Wang, J.J. Wen, W.G. Bornmann, C.M. Crews, The anti-angiogenic agent fumagillin covalently binds and inhibits ths methionine aminopeptidase MetAP-2, Proc. Natl. Acad. Sci. USA 12 (1997) 6099–6103.
[12] M.D. Katinka, S. Duprat, E. Cornillot, G. Metenier, F. Thomarat, G. Prensier, et al., Genome sequence and gene compaction of the eukaryote parasite *E. cuniculi*, Nature 414 (2001) 450–453.
[13] R. Li, X. Chen, B. Gong, P.M. Selzer, Z. Li, E. Davidson, et al., Structure-based design of parasitic protease inhibitors, Bioorg. Med. Chem. 4 (1996) 1421–1427.
[14] C.L. Byington, R.L. Dunbrack, F.G. Whitby, F.E. Cohen, N. Agabian, *Entamoeba histolytica*: computer-assisted modeling of phosphofructokinase for the prediction of broad-spectrum antiparasitic agents, Exp. Parasitol. 87 (1997) 194–202.
[15] S. Liu, J. Widom, C.W. Kemp, C.M. Crews, J. Clardy, Structure of human methionine aminopeptidase-2 complexed with fumagillin, Science 282 (1998) 1324–1326.
[16] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, J. Mol. Biol. 234 (1993) 779–815.
[17] R.A. Laskowsky, J.A.C. Rullmann, M.W. MacArthur, R. Kaptein, J.M. Thronton, AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR, J. Biomol. NMR 8 (1996) 477–486.
[18] A.T. Brünger, X-PLOR Version 3.1, a System for X-ray Crystallography and NMR, Yale University Press, New Haven and London, 1992.

[19] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, et al., GAUSSIAN 98, Revision A.9, Gaussian Inc, Pittsburgh, PA, USA, 1998.
[20] R. Koradi, M. Billeter, K. Wüthrich, MOLMOL: a program for display and analysis of macromolecular structures, J. Mol. Graphics 14 (1996) 51–55.
[21] W.T. Lowther, Y. Zhang, P.B. Sampson, J.F. Honek, B.W. Matthews, Insights into the mechanism of the *E. coli* methionine aminopeptidase from the structural analysis of the reaction products and phosphorus-based transition state analogues, Biochemistry 38 (1999) 14810–14819.
[22] T.H. Tahirov, H. Oki, T. Tsukihara, K. Ogashara, K. Yutani, K. Ogata, et al., Crystal structure of the methionine amminopeptidase from hyperthermophile *P. furiosus*, J. Mol. Biol. 284 (1998) 101–124.
[23] A.D. Becke, Density-functional exchange-energy approximation with correct asymptotic behavior, Phys. Rev. 38 (1988) 3098.
[24] P.J. Hay, W.R. Wadt, Ab-initio effective core potential for molecular calculation. Potential for K to Au including the outermost core orbital, J. Chem. Phys. 82 (1985) 299–310.
[25] A.E. Reed, L.A. Curtis, F. Weinhold, Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint, Chem. Rev. 88 (1988) 899–926.
[26] S.L. Roderick, B.W. Matthews, Structure of the cobalt-dependent methionine aminopeptidase from *E. coli*: a new type of proteolytic enzyme, Biochemistry 32 (1993) 3907–3912.
[27] W.T. Lowther, B.W. Matthews, Structure and function of the methionine aminopeptidase, Biochem. Biophys. Acta 1477 (2000) 157–167.