Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 25 (2006) 10–16
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Human cytidine deaminase: A three-dimensional homology model of a tetrameric metallo-enzyme inferred from the crystal structure of a distantly related dimeric homologue

Stefano Costanzi [a,*], Silvia Vincenzetti [b], Gloria Cristalli [c], Alberto Vita [b]

[a] *Computational Chemistry Core Laboratory, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, 12A Center Drive Rm 4051 MSC 5646, Bethesda, MD 20892-0810, USA*
[b] *Dipartimento di Scienze Veterinarie, University of Camerino, Camerino 62032, Italy*
[c] *Dipartimento di Scienze Chimiche, University of Camerino, Camerino 62032, Italy*

Received 11 July 2005; accepted 19 October 2005
Available online 21 November 2005

Abstract

Cytidine deaminase (CDA) is a cytosolic metalloprotein whose functional unit can be either a homotetramer (T-CDA) or a homodimer (D-CDA), depending on the species. In 1994, the first crystal structure of the dimeric *Escherichia coli* CDA has been published. However, a crystal structure of a tetrameric CDA was not determined until 2002. Prior to the disclosure of the experimentally elucidated structure of a tetrameric CDA, we derived a homology model of the human T-CDA employing the crystal structure of the dimeric *E. coli* CDA as a template. The comparison of our theoretical model with the crystal structure of the human T-CDA, subsequently published in 2004, validates our prediction: not only of the structural features of the monomer and the details of the binding site, but also the multimeric arrangement of the subunits were determined with high accuracy in our model. By means of a phylogenetic analysis conducted on CDAs from various organisms, we demonstrate that the *E. coli* CDA is one of the furthest known homologues of the human enzyme. Nonetheless, despite the evolutionary distance and, more importantly, the different multimeric arrangement of their functional units, the *E. coli* CDA proved to have all the necessary information to accurately infer the structure of its human homologue.

Keywords: Cytidine deaminase; Homology modeling; Sequence alignment; Phylogenetic analysis; Evolution

1. Introduction

Cytidine deaminase (CDA, EC 3.5.4.5) is a multimeric cytosolic metalloprotein, which catalyzes the formation of uridine and deoxyuridine by hydrolytic deamination of cytidine and deoxycytidine, respectively. Besides its natural substrates, CDA can also deaminate a number of antitumoral and antiviral cytidine analogs, such as cytosine arabinoside (AraC) and 5-aza-2′-deoxycytidine (5-Aza-CdR) [1], leading to their pharmacological inactivation.

Depending on the species, the CDA functional unit can be either a homotetramer (T-CDA) or a homodimer (D-CDA): the former is typical of most of the species including human [2,3], while the latter has been found in gamma-Proteobacteria, e.g. *Escherichia coli* [4,5], and flowering plants, e.g. *Arabidopsis thaliana* [6]. In both cases, an isolated CDA monomer completely lacks any catalytic activity.

In 1994, Betts et al. [7] published the first crystal structure of the *E. coli* D-CDA (PDB ID: 1CTU), which confirmed the homodimeric structure of the enzyme that we had previously hypothesized [4]. Each monomer, of a molecular weight of 31.5 kDa, was found to be composed of a small N-terminal alpha-helical domain with no evident function, a catalytic domain containing a zinc-binding pocket, and a C-terminal domain. The latter was found to be related to the catalytic domain by an approximate two-fold symmetry. The crystal structure also revealed that the dimer contains two active sites, located at the interface between the two subunits. A zinc ion, essential for the catalytic activity of the enzyme, was found in each active site, tetrahedrally coordinated by one His and two Cys residues of the catalytic domain, in addition to the O-4 hydroxyl group of the bound nucleoside inhibitor.

* Corresponding author. Tel.: +1 301 451 7353; fax: +1 301 480 4586.
E-mail address: stefanoc@mail.nih.gov (S. Costanzi).

1093-3263/$ – see front matter
doi:10.1016/j.jmgm.2005.10.008

In the following years, the same group solved additional structures of the *E. coli* D-CDA complexed with various ligands. However, until 2002, when the structure of the T-CDA from *Bacillus subtilis* (PDB ID: 1JTK) was published by Johansson et al. [8], the *E. coli* D-CDA has been the only experimentally elucidated CDA structure. Finally, in 2004 a crystal structure of the human enzyme (PDB ID: 1MQ0) has been published [9].

In the absence of experimental data on the structure of the human T-CDA, in an effort to interpret the results of our site-directed mutagenesis data and to rationally design future experiments [3,10], we relied on a homology model of the human T-CDA which we constructed on the basis of the structure of the *E. coli* D-CDA [11–13]. The *E. coli* and the human enzymes showed only about 27% sequence identity. Furthermore, our experimental evidences indicated that the functional unit of the human enzyme was a tetramer composed of four identical monomers of 14 kDa, which contained a total of four binding sites and four zinc ions. Despite the structural differences with the available template, our homology model appeared to be consistent with the experimental data.

In the present paper we validate the accuracy of our homology model of the human T-CDA by comparing it with the recently published crystal structure. Furthermore, we describe the strategy that we used to exploit the crystal structure of the dimeric *E. coli* D-CDA to infer the model of the tetrameric human T-CDA. We also report a phylogenetic analysis of CDAs from various species, aimed at putting into an evolutionary context the hereby presented comparative modeling, while clarifying the degree of similarity among the studied homologues.

2. Methodologies

2.1. Sequence analysis

The alignment of 113 sequences of CDA from various organisms was performed with the program CLUSTALX [14], the Windows interface of CLUSTALW [15]. The Dayhoff percentage of accepted mutation (PAM) substitution matrix [16] was used with a Gap Start penalty of 12 and a Gap Extend penalty of four. The list of the sequences, grouped by class, is provided as supporting information (Appendix A). Due to the significant difference in the length of the sequences between the T-CDAs and the D-CDAs, the multiple sequence alignment was performed in three steps. In the first step the 75 sequences of the T-CDAs were aligned. In the second step, the 38 remaining sequences of the D-CDAs were added and aligned to the others without allowing any change in the relative alignment of the T-CDAs obtained in the previous step. In the third step all the sequences were aligned together without the use of any constraint. A pairwise distance matrix was calculated on the sequence alignment by means of the Protdist program of the Phylip Package [17] using the Dayhoff PAM substitution matrix [16].

Subsequently, the CDA sequences were divided by class, and a consensus sequence was generated for each class. All the columns with gaps were stripped from the resulting consensus alignment, and a new PAM pairwise distance matrix was calculated. From the distance matrix, a rooted phylogram was reconstructed according to the UPGMA (unweighted pair group method with the arithmetic average) algorithm [18] as implemented in the program Neighbor of the Phylip Package [17].

2.2. Molecular modeling

Homology modeling, energy minimization, and molecular dynamics simulations were carried out utilizing the molecular operating environment (MOE) [19]. For molecular mechanics calculations, the AMBER 94 [20] force field was employed with a distance dependent dielectric constant. Energy minimization was performed employing the truncated Newton method.

Fig. 1 shows the pairwise alignment of the sequences of the human and *E. coli* CDAs which was used to construct the homology model of the human T-CDA monomer using the catalytic domain of the *E. coli* D-CDA complexed with uridine (1af2.pdb) [21] as a template.

The structure of the human homotetramer was inferred from the structure of the *E. coli* homodimer by exploiting the internal two-fold symmetry of the *E. coli* monomer, as explained in detail in Section 3. The structural superimpositions necessary to assemble the human T-CDA tetramer were performed by means of the superpose command of MOE [19], taking into account the backbone atoms only. We made use of the ‘‘accent secondary structure matches’’ feature, which increases the weights of residues with matching regular secondary structure.

A uridine molecule, a zinc ion, and a leaving ammonia molecule were positioned into the four active sites of the modeled human T-CDA by superimposing the catalytic domain of the *E. coli* enzyme to each of the four modeled human monomers. The zinc ions were then coordinated with the sulfur atoms of Cys65, Cys99 and Cys102 and to the uridine O-4.

The entire structure was submitted to a preliminary energy minimization, followed by 50 ps of molecular dynamic simulation at 300 K and by a final energy minimization until an RMS of 0.001 kcal/mol/Å was reached.

3. Results and discussion

3.1. Sequence analysis

Following the procedure described in above, we compared the sequences of 113 CDAs from different species by means of a multiple sequence alignment followed by a phylogenetic analysis. The distances, calculated according to the Dayhoff method, are expressed in percentage of accepted mutation units, which take into account the possibility of multiple substitutions at the same position. For example, a distance of 100 PAM indicates 100 mutations per 100 amino acids and due to the possibility of multiple substitutions, roughly corresponds to 43% observable differences [22]. In Fig. 2, we report a chart showing the average distance of the studied CDAs, grouped by

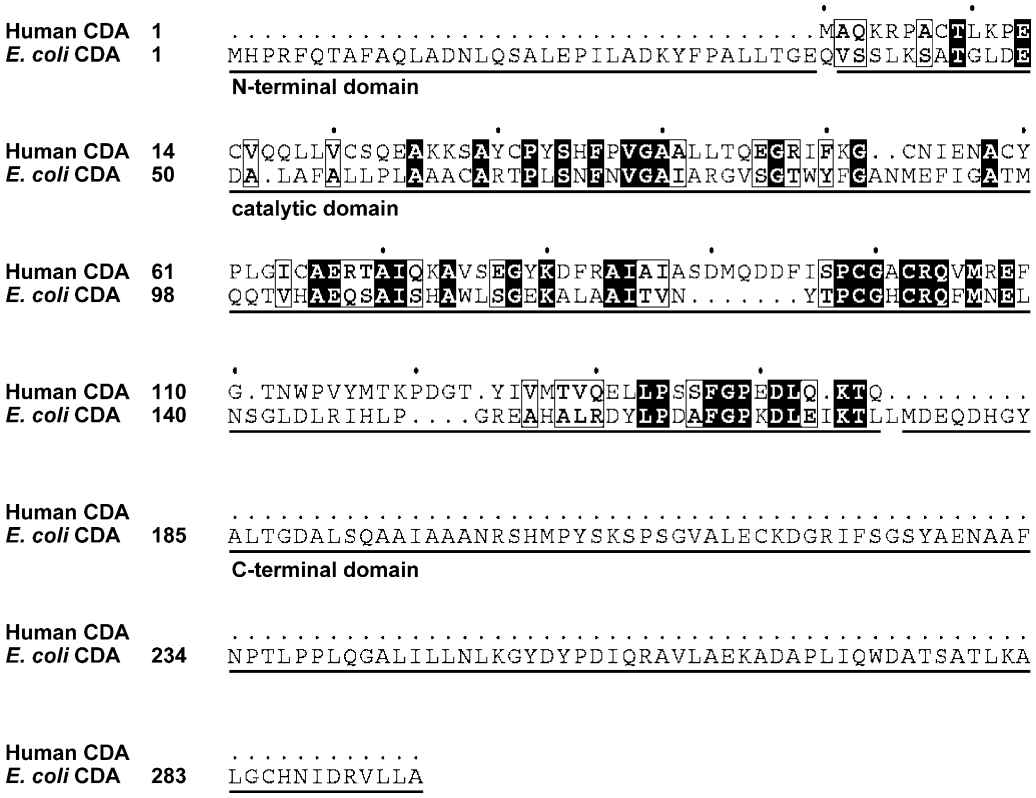

Fig. 1. Sequence alignment of the human T-CDA and the *E. coli* D-CDA monomers. The length of the *E. coli* subunit is over twice as long as the human one. Besides the central catalytic domain, every *E. coli* monomer contains an N-terminal domain and a C-terminal domain.

class, from the human enzyme. The list of the sequences grouped in each class is provided as supporting information. For every listed sequence, the calculated PAM distance from the human T-CDA is also reported. Not surprisingly, the T-CDAs from mouse, showing a distance of about 20 PAM, represent the closest homologues of the human enzyme. At a distance of about 100 PAM, the next closest homologues are the eukaryotic T-CDAs from Diplomonadida and, interestingly, the prokaryotic T-CDAs from Bacilli and Clostridia. The D-CDAs from gamma-Proteobacteria, e.g. *E. coli*, and Magnoliophyta are located at a significant distance from the human enzyme of 280 and 300 PAM, respectively. Based on this analysis, the *E.*

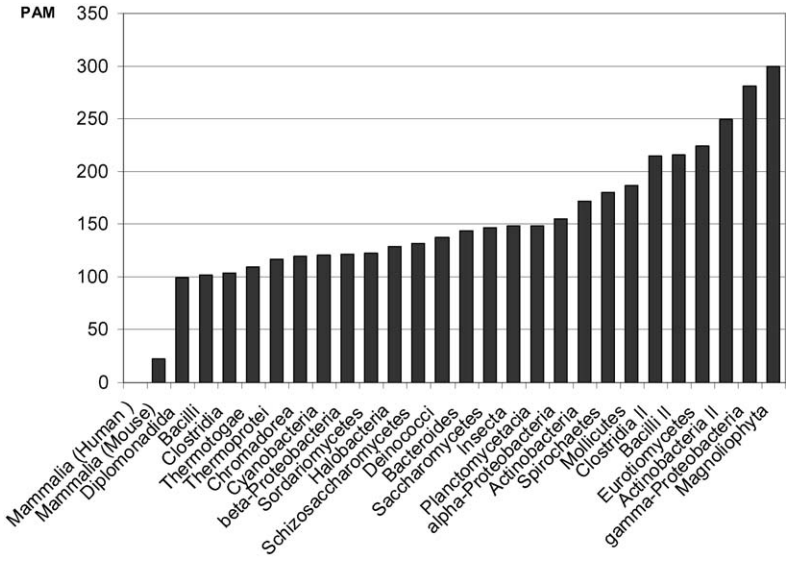

Fig. 2. Average percentage of accepted mutation (PAM) distances among the analyzed CDAs, divided by class of organism, from their human homologue. To facilitate the comparison the mammalian T-CDAs from human and mouse are listed separately. The calculation takes into account the possibility of multiple substitutions at the same site. The D-CDAs from gamma-Proteobacteria, e. g. *E. coli*, are among the furthest homologues of the human enzyme, showing an average distance of about 280 PAM.

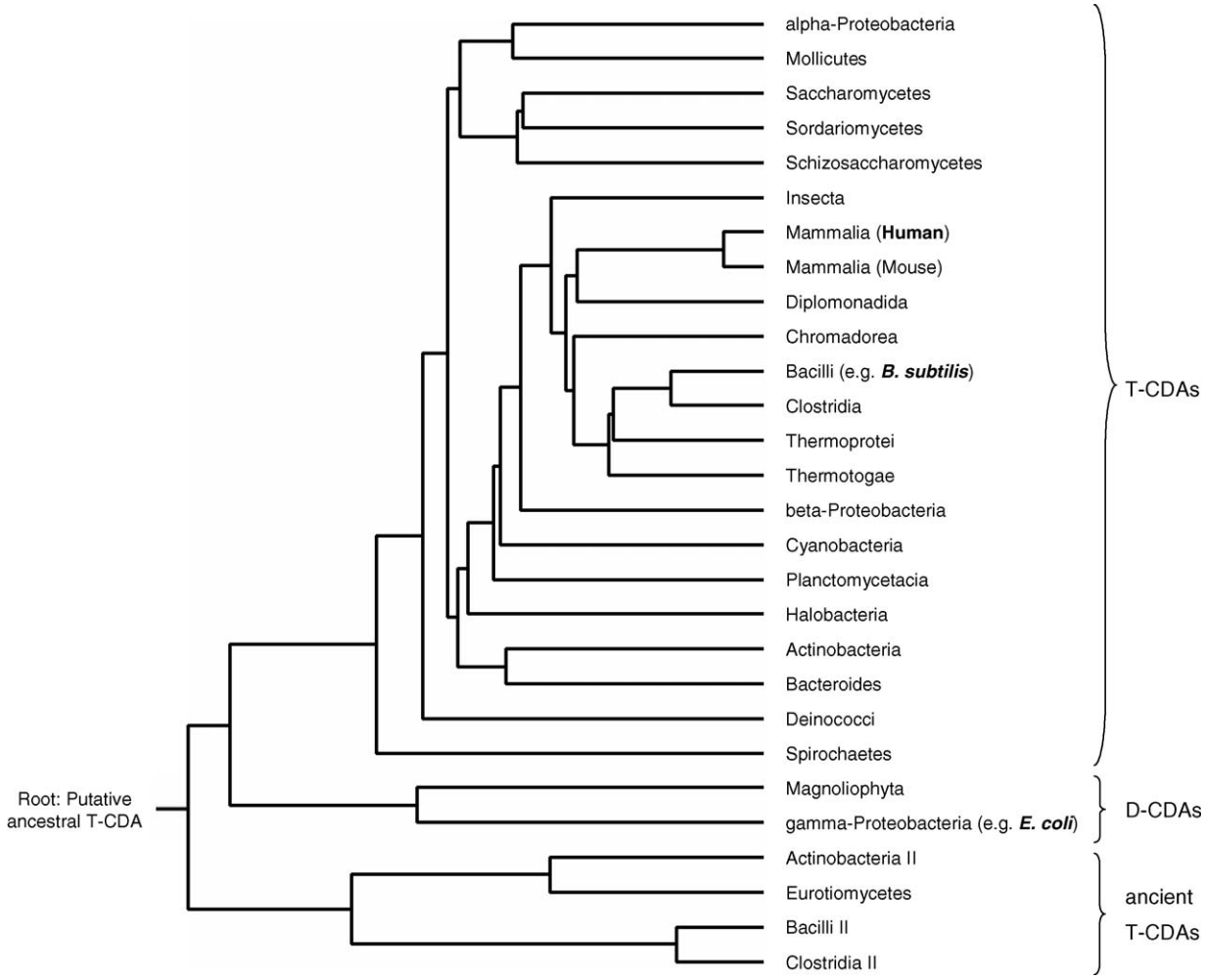

Fig. 3. Rooted UPMGA phylogram reflecting the hypothetical evolutionary relationships among the studied CDAs, divided by class of species. To facilitate the comparison the mammalian T-CDAs from human and mouse are listed separately. The three organisms for which the three-dimensional structure of the enzyme has been determined by X-ray crystallography are shown in bold. The homotetramer is the functional unit for all the reported CDAs, with the exceptions of the enzymes from magnoliophyta and gamma-Proteobacteria, which belong to the family of the homodimeric CDAs (D-CDAs). The horizontal length of the branches is proportional to the PAM distances among the sequences.

coli D-CDA, which was used as template for the modeling described in this paper, is one of the furthest homologous of the human T-CDA.

Fig. 3 shows a rooted UPMGA [18] phylogram, calculated on the PAM distance matrix as explained above, which represents the putative evolution of the sequences from a common ancestor. The great majority of the T-CDAs, including those from human and mouse, are clustered together in the phylogram. However, Bacilli, Clostridia, and Actinobacteria, besides the isoforms included in the T-CDA cluster, show also a second, more ancient, isoform of T-CDA clustered together with the T-CDA from Eurotiomycetes. Among all the analyzed CDAs, these appear to be the first to have differentiated from the other homologues. This observation suggests that T-CDAs and D-CDAs descend from a common homotetrameric ancestor from which, at some point in the evolution, the D-CDAs originated. Besides the diverging T-CDAs from the four aforementioned classes of species, the D-CDAs showed the highest evolutionary distance from the human T-CDA.

3.2. Inferring the three-dimensional structure of a homotetrameric protein using a dimer as template

In addition to being distantly related to its human homologue, the *E. coli* enzyme is endowed with a different multimeric structure. Therefore, once we built the homology model of the human T-CDA monomer using the catalytic domain of the *E. coli* D-CDA as template, we faced the problem of assembling the tetrameric structure of the human enzyme.

Since the *E. coli* CDA monomer exhibits an approximate two-fold symmetry between the catalytic and the C-terminal domains and since it weights twice as much as the human monomer, we hypothesized that a pair of monomers of the human homotetramer could be arranged in a way that resembles the structure of one monomer of the *E. coli* dimer. Therefore, we devised the strategy schematized in Fig. 4. Following this scheme, we divided the *E. coli* dimer into four parts: A1 (catalytic domain of chain A), A2 (C-terminal domain of chain A), B1 (catalytic domain of chain B), and B2 (C-terminal

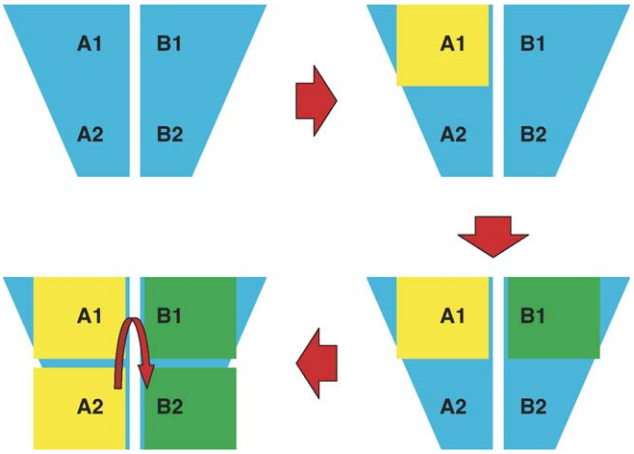

Fig. 4. Schematic representation of strategy employed for the assembly of the T-CDA homotetrameric structure. The *E. coli* dimer, represented in cyan, has been divided in four parts: A1 (catalytic domain of chain A), A2 (C-terminal domain of chain A), B1 (catalytic domain of chain B), and B2 (C-terminal domain of chain B). A dimer of the human enzyme was constructed by superimposing a first human monomer (yellow) to the *E. coli* catalytic domain A1, and a second one to (green) and to the catalytic domain B1. Subsequently, the generated human dimer has been replicated and rigidly superimposed onto the C-terminal domains A2 and B2 of the *E. coli* dimer to obtain the functional tetrameric structure.

domain of chain B). A dimer of the human enzyme was constructed by superimposing a first human monomer to the *E. coli* catalytic domain A1, and a second one to the catalytic domain B1. The resulting dimer was then replicated and rigidly superimposed on the *E. coli* C-terminal domains A2 and B2 to give a homotetrameric structure showing an overall C222 symmetry (Fig. 5). The rigid superimposition of a pre-constructed human dimer onto the C-terminal domains of the *E. coli* dimer has been the key step in assembling the multimeric structure of the human enzyme that later was experimentally validated as described below. In fact, the limited structural correspondence between the human CDA monomer and the C-terminal domain of the *E. coli* enzyme would have made impossible to build the tetrameric structure by simply superimposing each individual human monomer onto each individual *E. coli* C-terminal domain. Fig. 5 shows a superimposition of our homology model of the human T-CDA and the crystal structure of the *E. coli* D-CDA. The overall three-dimensional structure of the *E. coli* homodimer is well resembled by the human homotetramer. Nonetheless, the two enzymes show substantial differences: the N-terminal domains of the *E. coli* D-CDA have no correspondence in the human T-CDA, while the C-terminal domains of the *E. coli* D-CDA are considerably smaller than the corresponding human T-CDA monomers.

After the publication of the crystal structure of the homotetrameric *B. subtilis* T-CDA [8] we were able to compare our homology model to a more strictly related protein. Finally, with the publication of the crystal structure of the human enzyme itself [9], we directly validated the accuracy of our theoretical model by calculating the root mean square deviation (RMSD) between the backbone atoms of the crystal structure and our model. After the superimposition of the single monomers, it amounted to 3.36 Å. The highest divergence was found in the domains 13–22 (RMSD 5.38 Å) and 119–127 (RMSD = 8.43 Å). When these two domains were excluded from the calculation, the RMSD decreased to 2.46 Å. Remarkably, the RMSD of the backbone atoms calculated on the whole tetrameric structure was found to be 3.50 Å, which decreased to 2.58 Å when the two domains showing the highest divergence were excluded from the calculation. Hence, the multimeric structure of our human T-CDA theoretical model resulted in very good agreement with the experimentally elucidated structure. These results are well illustrated in Fig. 6,

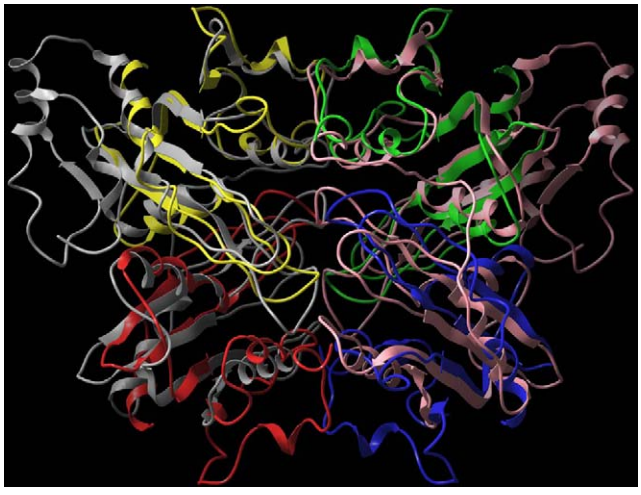

Fig. 5. Ribbon representation of the human T-CDA homology model superimposed to the *E. coli* D-CDA crystal structure. The two monomers of the *E. coli* enzyme are colored in gray and pink, while the monomers of the human enzyme are colored in yellow (A1), red (A2), green (B1), and blue (B2).

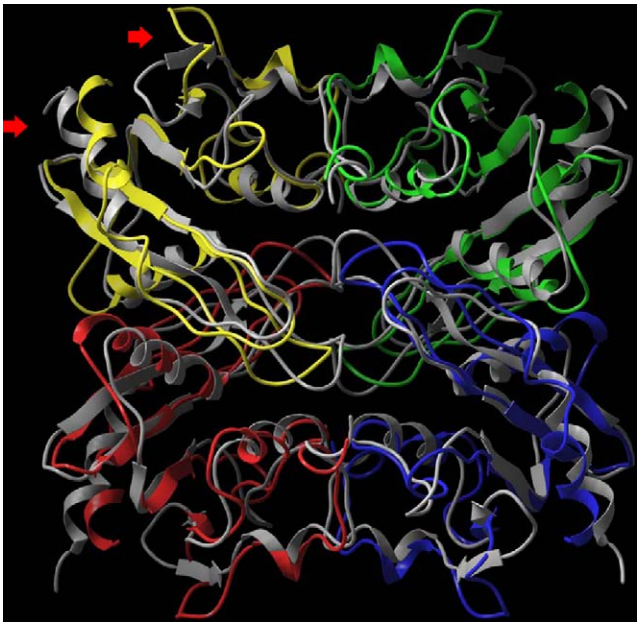

Fig. 6. Ribbon representation of human T-CDA homology model superimposed to the human T-CDA crystal structure. The crystal structure is colored in gray, while the monomers of the human enzyme are colored in yellow (A1), red (A2), green (B1), and blue (B2). The red arrows indicate the two domains showing the highest divergence between the model and the crystal structure.

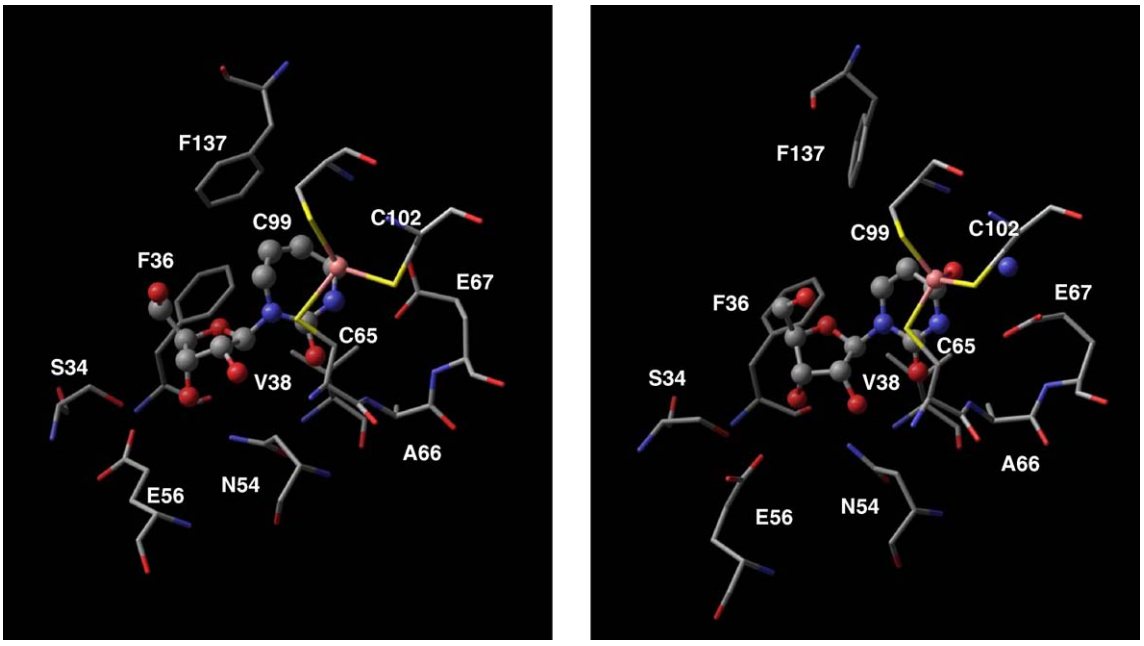

Fig. 7. Comparison of the active sites of the human T-CDA crystal structure complexed with the inhibitor diazepinone riboside (left) and the human T-CDA homology model complexed with uridine and ammonium (right). All the amino acids belong to the A1 monomer with the exception of Phe137, which belongs to the B1 monomer. The amino acids are represented in sticks whiles the nucleosides, the ammonium, and the zinc ion are represented in balls and sticks.

which shows a superimposition of our homology model of the human T-CDA and the crystal structure of the same enzyme. Even if distantly related from the human homologue and endowed with a different multimeric functional unit, the *E. coli* dimer proved to have all the information necessary to infer an accurate model of the human tetramer.

The *E. coli* D-CDA proved to be a good template to model the human enzyme also with regard to the topology of the active site (Fig. 7). We suggested that the zinc ion would be tetrahedrally coordinated by the uridine O-4 and the thiolate side chains of Cys65, Cys99, and Cys102 [3,11,12]. We suggested that Glu 67, which plays a central role in the mechanism of the catalysis as explained elsewhere [12], would be located in proximity of the zinc ion and the 4 and 3 positions of the pyrimidine ring of the bound nucleoside [3,11,12]. We identified Asn 54 and Glu 56 as responsible of the coordination of the uridine OH-2′ and OH-3′ [11,12]. In addition, we suggested that the amino acids of three different subunits concurred to form the binding pocket for the substrate [10–13]. In particular Phe137 of subunit B1, was suggested to be located at the interface between the subunits A1 and B1 and to concur to the formation of the binding pocket for the ligand pyrimidine moiety. The particular location of Phe137 explained the contribution given by this residue to the stabilization of the CDA quaternary structure that we demonstrated experimentally [10]. All these structural hypotheses inferred by our model were validated by the crystal structure of the human T-CDA.

4. Conclusions

Within the CDA family, the human and the *E. coli* enzymes are among the furthest homologues. Indeed, the sequence of the *E. coli* enzyme showed a distance of about 280 PAM from the human enzyme, meaning that as many as 280 substitutions per 100 residues may have taken place during the course of the evolution. Due to the occurrence of multiple substitutions at the same site, the observed non-conserved residues between the two enzymes were about 73% of the entire human sequence.

Despite the evolutionary distance and, more importantly, the different multimeric arrangement of their functional units, the *E. coli* CDA proved to have all the necessary information to infer the structure of its human homologue. In this work, we demonstrated the possibility of deriving an accurate model of the human T-CDA on the basis of the dimeric structure of the *E. coli* D-CDA: not only the structural features of the monomer and the details of the binding site, but also the three-dimensional arrangement of the subunits in the biologically functional homotetramer were determined with high precision in our model.

In conclusion, the comparison of our homology model of the human T-CDA with its crystal structure strongly supports the theory according to which, during the evolution of a family of proteins from a common ancestor, the conservation of the three-dimensional structure is higher than the conservation of the sequences. In other words, homologous proteins, i.e. proteins with the same biological function and with a common evolutionary origin, are generally endowed with the same three-dimensional features.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2005.10.008.

References

[1] J. Lalibertè, V.E. Marquez, L.R. Momparler, Potent inhibitors for the deamination of cytosine arabinoside and 5-aza-2′-deoxycytidine by human cytidine deaminase, Cancer Chemother. Pharmacol. 30 (1992) 7–11.

[2] S. Vincenzetti, A. Cambi, J. Neuhard, E. Garattini, A. Vita, Recombinant human cytidine deaminase: expression, purification and characterization, Protein Expr. Purif. 8 (1996) 247–253.

[3] A. Cambi, S. Vincenzetti, J. Neuhard, S. Costanzi, P. Natalini, A. Vita, Identification of four amino acid residues essential for catalysis in human cytidine deaminase by site-directed mutagenesis and chemical modification, Protein Eng. 11 (1998) 59–63.

[4] A. Vita, A. Amici, T. Cacciamani, M. Lanciotti, G. Magni, Cytidine deaminase from *Escherichia coli* B. Purification and enzymatic and molecular properties, Biochemistry 24 (1985) 6020–6024.

[5] C. Yang, D. Carlow, R. Wolfenden, S.A. Short, Cloning and nucleotide sequence of the *Escherichia coli* cytidine deaminase (cdd) gene, Biochemistry 31 (1992) 4168–4174.

[6] S. Vincenzetti, A. Cambi, J. Neuhard, K. Schnorr, M. Grelloni, A. Vita, Cloning, expression, and purification of cytidine deaminase from *Arabidopsis thaliana*, Protein Expr. Purif. 15 (1999) 8–15.

[7] L. Betts, S. Xiang, S.A. Short, R. Wolfenden, C.W. Carter, Cytidine deaminase. The 2.3 Å crystal structure of an enzyme: transition-state analog complex, J. Mol. Biol. 235 (1994) 635–656.

[8] E. Johansson, N. Mejede, J. Neuhard, S. Larsen, Crystal structure of the tetrameric cytidine deaminase from *Bacillus subtilis* at 2.0 Å resolution, Biochemistry 41 (2002) 2563–2570.

[9] S.J. Chung, J.C. Fromme, G.L. Verdine, Structure of human cytidine deaminase bound to a potent inhibitor, J. Med. Chem. 48 (2005) 658–660.

[10] S. Vincenzetti, G. De Sanctis, S. Costanzi, G. Cristalli, P. Mariani, G. Mei, J. Neuhard, P. Natalini, A. Vita, Functional properties of subunit interactions in human cytidine deaminase, Protein Eng. 16 (2003) 1055–1061.

[11] S. Costanzi, S. Vincenzetti, A. Vita, C. Lambertucci, S. Taffi, S. Vittori, R. Volpini, G. Cristalli, Human cytidine deaminase: structural analysis of an homotetrameric metalloprotein through homology modelling, in: Molecular Graphics and Modelling Society Annual Iinternational Meeting, Bristol, UK, April 2–5, 2002.

[12] S. Costanzi, S. Vincenzetti, A. Vita, C. Lambertucci, S. Taffi, R. Volpini, S. Vittori, G. Cristalli, Human cytidine deaminase: understanding the catalytic mechanism, Nucleosides Nucleotides Nucleic Acids 22 (2003) 1539–1543.

[13] S. Vincenzetti, S. Costanzi, G. Cristalli, P. Mariani, B. Quadrini, N. Cammertoni, A. Vita, Intersubunit interactions in human cytidine deaminase, Nucleosides Nucleotides Nucleic Acids 22 (2003) 1535–1538.

[14] J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, D.G. Higgins, The CLUSTAL X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, Nucleic Acids Res. 25 (1997) 4876–4882.

[15] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673–4680.

[16] M.O. Dayhoff, Atlas of Protein Sequence and Structure, Suppl. 3, vol. 5, National Biomedical Research Foundation, Washington, DC, 1978, pp. 345–352.

[17] PHYLIP Phylogeny Inference Package, Version 3.6, 2002, Joseph Felsenstein, Department of Genome Sciences, University of Washington, Seattle, WA, USA.

[18] R.R. Sokal, C.D. Michener, A statistical method for evaluating systematic relationships, University of Kansas Scientific Bulletin 28 (1958) 1409–1438.

[19] MOE, Version 2001.01, 2001, Chemical Computing Group, Inc., Montreal, Canada.

[20] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz Jr., D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, J. Am. Chem. Soc. 117 (1995) 5179–5197.

[21] S. Xiang, S.A. Short, R. Wolfenden, C.W. Carter, The structure of the cytidine deaminase-product complex provides evidence for efficient proton transfer and ground-state destabilization, Biochemistry 36 (1996) 4768–4774.

[22] H.B. Nicholas, D.W. Deerfield, A.J. Ropelewski, Strategies for searching sequence databases, Biotechniques 28 (2000) 1174–1191.