BIOORGANIC CHEMISTRY 23, 427–438 (1995)

Modeling the Three-Dimensional Structure of Serpin/Molecular Chaperone HSP47

JOSEPH W. DAVIDS,* TALAL S. H. EL-THAHER,* AKIRA NAKAI,† KAZUHIRO NAGATA,† AND ANDREW D. MILLER,*,[1]

*Imperial College of Science, Technology & Medicine, Department of Chemistry, South Kensington, London, SW7 2AY, United Kingdom; and †Department of Cell Biology, Chest Disease Research Institute, Kyoto University, Kyoto 606, Japan

Received May 12, 1995

Heat shock protein 47 (HSP47) is a molecular chaperone which assists procollagen triple helix assembly and secretion. As a molecular chaperone it is unique in that it binds specifically to a very narrow range of protein "substrates" (i.e., procollagen and collagen only), and it is also a member of the well-characterized serine protease inhibitor (serpin) superfamily. In the absence of any X-ray crystallographic data, a novel tandem-modeling procedure is used to obtain three-dimensional structural information on mature recombinant mouse HSP47 (mrmHSP47). MrmHSP47 is shown to have 30% amino acid sequence identity and 70% sequence similarity with human protein C inhibitor (hPCI). Therefore, molecular models of inhibitory and latent states of hPCI are generated, using the X-ray crystal structure coordinates of proteolytically cleaved hPCI, and used as templates for the homology modeling of inhibitory and latent states of mrmHSP47. The validity of the models is discussed and the latent state model of mrmHSP47 is shown to have a suitable candidate binding groove for procollagen/collagen peptides, which appears to account for experimental observations made with amino acid deletion experiments. © 1995 Academic Press, Inc.

INTRODUCTION

Molecular chaperones are a ubiquitous, abundant, and highly conserved group of proteins which assist protein folding/refolding *in vitro* and *in vivo* (*1*, *2*) as well as protecting proteins from stress-induced unfolding (*1*). They first came to attention because of their specific induction during the cellular response of all organisms to heat shock (*3*, *4*) but are now known to be constitutively and abundantly expressed in the absence of any stress. Heat shock protein 47 (HSP47)[2] is one such molecular chaperone. While the vast majority of molecular chaperones interact with a large variety of globular proteins (*1*, *2*, *5*, *6*), HSP47 (also known as colligin (*7*) or J6 protein (*8*) is specific only to polypeptides of procollagen and the fibrous protein collagen. There is now considerable evidence that HSP47 is a molecular chaperone for the formation of procollagen triple helices and also for the secretion of procolla-

[1] To whom correspondence should be addressed. Fax: (44) (171) 594 5803.

[2] Abbreviations used: HSP47, heat shock protein 47; mrmHSP47, mature recombinant mouse HSP47; serpin, serine protease inhibitor; hPCI, human protein C inhibitor; ACHY, antichymotrypsin, PAI-1, plasminogen activator inhibitor-1; AT III, antithrombin III.

0045-2068/95 $12.00
Copyright © 1995 by Academic Press, Inc.
All rights of reproduction in any form reserved.

gen from endoplasmic reticulum to the Golgi (*9–11*). Both bone and connective tissue are ultimately formed from these procollagen triple helices.

In order to understand the underlying mechanism of HSP47-assisted formation of procollagen triple helix, knowledge about the three-dimensional structure of this protein is essential. However, there is no X-ray crystal structure of HSP47 currently available which led us to consider an alternative molecular modeling approach to obtaining structural information with which to understand structure/function relationships in this unique molecular chaperone. Molecular modeling appeared to be a valid approach to obtaining realistic structural information since in addition to being a functional molecular chaperone, HSP47 is known (*9, 10, 12*) to be a member of the serpin (serine protease inhibitor) superfamily of proteins, which is one of the most widely and best structurally characterized protein families. There are at least 30 serpin proteins, including HSP47, which have been identified on the basis of their primary amino acid sequences (*12, 13*) and where X-ray crystal structure information is available the three-dimensional structures are surprisingly consistent (*13, 14*). As a result, it is not unreasonable that HSP47 should adopt a similar protein fold and three-dimensional structure. Not all these serpins are functional serine protease inhibitors, but many do contain an ~28-amino residue loop (serpin loop) which is responsible for specific protease inhibition. Binding of this loop to a specific serine protease is followed by eventual dissociation of the serpin in a cleaved form in which the amino acids to the N- and C-terminal sides of the cleaved, scissile peptide bond (i.e., respectively, the P1 and P1′ amino acid residues according to the Pn nomenclature of Schechter and Berger (*15*) become separated by about 70 Å. Indeed, many of the X-ray crystal structures of serpins which have been determined (*14*) are of such proteolytically cleaved serpins. Exceptions to this are the recent X-ray crystal structures of uncleaved ovalbumin (*16*), antichymotrypsin (ACHY) (*17*), plasminogen activator inhibitor-1 (PAI-1) (*18*), and antithrombin III (AT III) (*19, 20*). These recent structures have revealed that uncleaved serpins may adopt one of two main conformations. In one, the uncleaved serpin loop projects out of the protein in a conformation accessible to a serine protease active site. This inhibitory state is typified by a conformation observed in the X-ray crystal structure of AT III (*19, 20*). In the other, the serpin loop does not protrude, but is rendered inaccessible to a serine protease active site. This so-called latent, noninhibitory state is typified by the X-ray crystal structure of PAI-1 (*18*) and an alternative conformation of AT III (*19*).

As a result of this wealth of structural data and consistency within the serpin superfamily, several studies have been published (*21–23*) in which molecular modeling techniques have been used to successfully predict the structures of serpin proteins, for which no X-ray crystal structures were then available, using the X-ray crystal structure coordinates of other serpin family members. Therefore, in the light of this literature precedent, the following paper outlines the application of molecular modeling, including the homology model approach, to derive appropriate three-dimensional structural information on serpin/molecular chaperone HSP47. Recently, the gene for mature recombinant mouse HSP47 (mrmHSP47) was cloned and overexpressed in *Escherichia coli* (*9*); therefore, we elected to model this protein specifically.

METHODS

Human Protein C Inhibitor Coordinates

The X-ray crystal structure coordinates of mature, cleaved human protein C inhibitor (hPCI) were available from the Brookhaven Protein Data Bank under the code 2pai. The structure begins at amino acid residue 5.

Sequence Alignment of Human Protein C Inhibitor and Mature Recombinant Mouse HSP47

The gene-derived amino acid sequence of mrmHSP47 (*9*) was optimally aligned with the amino acid sequence of hPCI using an initial conventional pairwise sequence alignment involving the Needleman–Wunsch algorithm (*24*), followed by sequence core analysis using the Dayhoff mutation (PAM250) matrix (*25*), and finishing with a review of residue hydrophobicities (*26*).

Molecular Modeling of Human Protein C Inhibitor

Model structures of uncleaved hPCI were generated from the known X-ray crystal structure coordinates (2pai) on a Silicon Graphics Iris Model 3000 workstation using the molecular modeling facilities of the protein imaging software package QUANTA 3.3. Two models of uncleaved hPCI were generated corresponding to a putative inhibitory and a putative latent, noninhibitory state, respectively. The putative inhibitory state of hPCI was modeled by rotating the ϕ and ψ backbone dihedral angles of both P10 and P12 (residues Ala349 and Ala347) in order to bring the P1 and P1′ residues (Arg358 and Ser359) into close proximity. A peptide bond was then created between P1 and P1′ residues, after which the structure was subject to energy minimization using CHARMm (*27*) to assist in the packing of side-chain atoms while harmonically constraining the α-carbon backbone. The putative latent state of hPCI was modeled by rotating the ϕ and ψ backbone dihedral angles of P12′ (residue Asn370) in order to bring the P1 and P1′ residues (Arg358 and Ser359) into close spatial proximity once more. Once again, a peptide bone was then created between P1 and P1′ residues, after which the structure was subject to energy minimization using CHARMm.

Homology Modeling of Mature Recombinant Mouse HSP47

Two structural models of mrmHSP47 were generated on a Silicon Graphics Iris Model 3000 workstation using the homology modeling facilities of the protein imaging software package QUANTA 3.3. A putative inhibitory state homology model of mrmHSP47 was generated using the inhibitory state model of hPCI (described above) as a template. First, the α-carbon backbone coordinates of inhibitory hPCI were assigned to corresponding residues of mrmHSP47, which had been found to correspond through the sequence alignment (see above). Undefined mrmHSP47 amino acid side-chain coordinates were then generated with idealized geometries using CHARMm. Subsequently, energy minimization, using a "Steepest Descents" algorithm (100 cycles) followed by an "Adopted Basis Set Newton

Raphson" algorithm (100 iterations), gave the completed putative inhibitory state homology model of mrmHSP47. A putative latent state homology model of mrmHSP47 was also generated by homology modeling in the same way using the latent state model of hPCI (described above) as a template.

Serine Protease Inhibition Studies with Mature Recombinant Mouse HSP47

MrmHSP47 was purified from recombinant *Escherichia coli* DH1λind-/pKS26 according to previously published protocols (*28*) and protein concentrations were estimated by the A_{280}/A_{260} absorption method of Warburg and Christian (*29*). Standard assays (*30, 31*) for thrombin and elastase were performed in the presence of mrmHSP47 at pH 7.5–7.8 using human thrombin (5 μg) and porcine pancreatic elastase (2 μM). The highest mrmHSP47 concentrations used were 38 and 51 μM, respectively. Standard trypsin, kallikrein, and tonin assays (*32*) were carried out in the presence of mrmHSP47 at pH 8.0 with bovine trypsin (100 ng), rat submandibular kallikrein (25 ng), and rat submandibular tonin (1.5 μg). The highest mrmHSP47 concentration used was 26 μM in all cases.

RESULTS

Sequence Alignment

MrmHSP47 is composed of a sequence of 400 amino acids with a molecular mass of 46,589 Da (*9*). Figure 1 shows the optimal sequence alignment of mrmHSP47 with hPCI. The sequence identity between the two proteins was 30% and the similarity 70% over 393 aligned residues. Figure 1 also indicates the serpin loop residues of hPCI and their equivalent in mrmHSP47 (from residues P12 through to P12') using the Pn nomenclature of Schechter and Berger (*15*).

Molecular Modeling of Human Protein C Inhibitor

The X-ray crystal structure coordinates of hPCI in a cleaved form were readily available from the Brookhaven Protein Data Bank. This structure (Fig. 2a) has been orientated and colored to explain the molecular modeling procedures by which models of inhibitory and latent states of hPCI were generated. Owing to the fact that all the currently solved three-dimensional X-ray crystal structures of serpins have the same basic secondary structural features (*13, 14*), Huber and Carrell (*13*) devised a careful labeling system so as to identify each of the secondary structural components individually (i.e., there are three β-sheets labeled A, B, and C with each strand separately numbered, e.g., s4A is strand 4 of β-sheet A; s1C is strand 1 of β-sheet C). The molecular modeling procedures revolved around β-sheet A (in yellow, Fig. 2a), strand s4A (in red, Fig. 2a), and strand s1C (in green, Fig. 2a). The inhibitory state of hPCI was modeled by perturbing the ϕ and ψ backbone dihedral angles of both residues P10 and P12 (Ala349 and Ala347) so that the P1 residue (Arg358) at the free end (carboxyl terminus) of β-strand s4A (in red, Fig. 2a) could be lifted out from β-sheet A and brought into close proximity with the

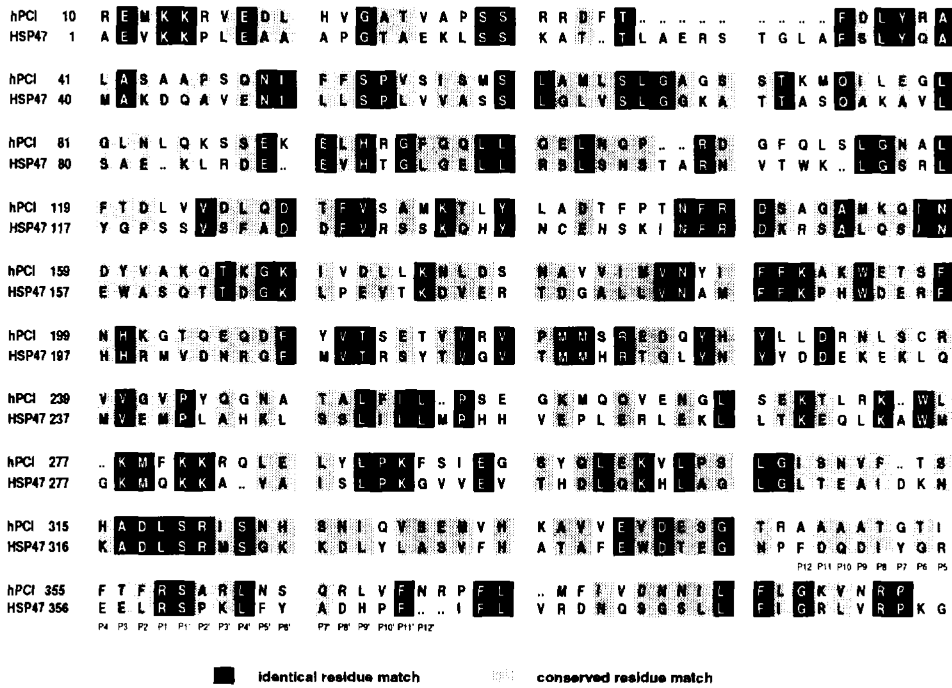
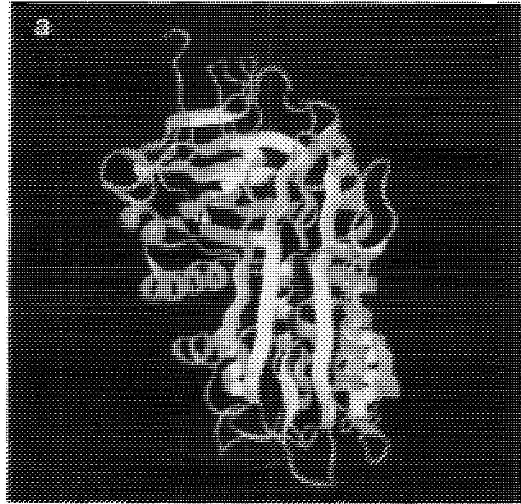
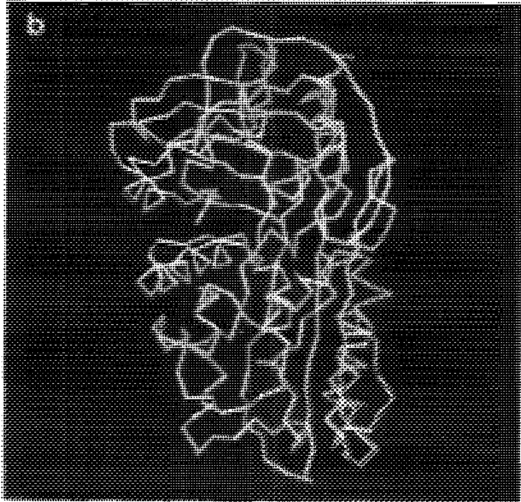
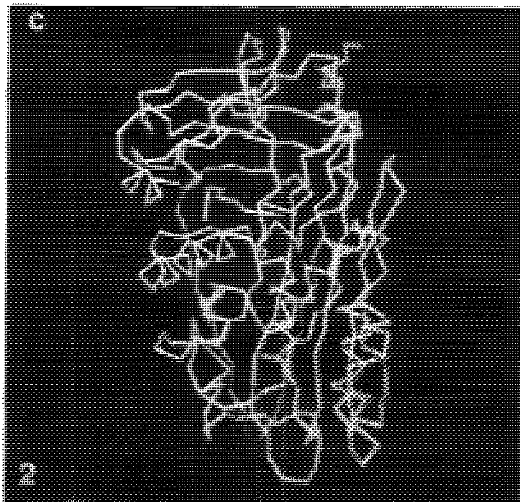


FIG. 1. Amino acid sequence alignment of mature recombinant mouse HSP47 (mrmHSP47) with human protein C inhibitor (hPCI). The serine protease inhibitor loop region of hPCI, and the corresponding region of mrmHSP47, is indicated using the Pn nomenclature of Schechter and Berger (*15*). Identical residues are shown against a black background and similar residues against a stippled background.

P1′ residue (ser359) at the free end (amino terminus) of strand s1C (in green, Fig. 2a). After the creation of a peptide bond between P1 and P1′ residues, energy minimization was used to optimize the inhibitory state model of hPCI (Fig. 2b) with the serpin loop (in yellow, Fig. 2b) projecting away from the remainder of the protein, thereby adopting a conformation accessible to a serine protease active site. The latent state model of hPCI was modeled by perturbing the ϕ and ψ backbone dihedral angles of residue P12′ (Asn370) only, so that the P1′ residue (ser 359) at the free end (amino terminus) of strand s1C (in green, Fig. 2a) could be brought into close proximity with the P1 residue (Arg358) at the free end (carboxyl terminus) of β-strand s4A (in red, Fig. 2a). This process involved the introduction of a little disruption to β-sheet C, while maintaining the integrity of β-sheet A. Once more, after the creation of a peptide bond between P1 and P1′ residues, energy minimization was used to optimize the latent state model of hPCI (Fig. 2c) with the serpin loop (in red, Fig. 2c) closely associated with the remainder of the protein thereby rendered inaccessible to a serine protease active site. Both inhibitory and latent state models of hPCI were superimposed on the X-ray crystal structure of the cleaved form of hPCI and the R.M.S. deviations recorded (Table 1).


a


b


c
2