Domain Flexibility in Aspartic Proteinases

Andrzej Sali, B. Veerapanandi, Jon B. Cooper, David S. Moss, Theo Hofmann, and Tom L. Blundell
ICRF Unit of Structural Molecular Biology, Department of Crystallography, Birbeck College, University of London, London WC1E 6HX, England; and Department of Biochemistry, King's College Circle, The University of Toronto, Toronto M5S 1A8, Canada

ABSTRACT

Comparison of the three-dimensional structures of native endothiapep- 
sin (EC 3.4.23.6) and 15 endothiopepsin oll- 
gopeptide inhibitor complexes defined at high resolution by X-ray crystallography shows that endothiopepsin exists in two forms differing in the relative orientation of a domain comprising residues 190–202. There are relatively few inter-
actions between the two parts of the enzyme; consequently, they can move as separate rigid bodies. A translational, librational, and screw analysis of the thermal parameters of endothi-
apepsin also supports a model in which the two parts can move relative to each other. In the comparison of different aspartic proteinases, the rms values are reduced by up to 47% when the two parts of the structure are superposed independently. This justifies description of the differences, including those between pepsi-

pepsin and pepsin (EC 3.4.34.1), as a rigid move-
ment of one part relative to another although considerable distortions within the domains also occur. The consequence of the rigid body movement is a change in the shape of the active site cleft that is largest around the $S_8$ pocket. This is associated with a different position and conformation of the inhibitors that are bound to the two endothiapepsin forms. The relevance of these observations to a model of the hydro-
lisis by aspartic proteinases is briefly dis-
cussed.

Key words: X-ray structure, TLS analysis, as-
partic proteinases, inhibitor com-
plexes, catalysis

INTRODUCTION

Aspartic proteinases are a class of proteolytic enzymes characterized by two essential aspartic res-

idues and specific inhibition by the microbial oli-
gopeptide papain. High-resolution X-ray struc-
tures of three fungal (EC 3.4.23.6) and three mammalian (EC 3.4.23.1) aspartic proteinases are available; the fungal enzymes include endo-

thiopepsin, 2, 3 penicillopepsin, 4, 5 and rhizopuspepsin, 6, 7 the mammalian enzymes are papain (EC 3.4.23.1), which has been refined in the monolucine 8, 9 and hexagonal 9 crystal forms, chymotry-

psin (EC 3.4.22.4), pepstegen, 10, 11, 12 and human renin. 13 Con-
dinates of several complexes between aspartic pro-
teinases and their oligopeptide inhibitors are also available. These include the structure of the com-
plex of rhinopuspepsin with its inhibitor 14 and 15 structures of endothiopepsin complexes, which have been solved at Birbeck College (Table I for names and references).

The fold of aspartic proteinases consists of two struc-
turally similar lobes of about 160 residues. 16 Each of these two lobes contributes one aspartate residue to the active site center, which is buried in the middle of a long and deep active site cleft at the interface between the two lobes. The approximate diad axis of symmetry relating the two lobes passes through the active site center. The fold consists pre-
dominantly of $B$-strands, but it also contains several helices. A structural comparison of mammalian and fungal aspartic proteinases and different forms of the same enzyme shows that, although the same sec-

ondary structure elements are used in all cases, the spatial relationships between some of them are con-
siderably different. See ref. 22 for the latest review of the structure and function of the aspartic protei-

nases.

Extensive data on specificity, steady- and pre-
steady-state kinetics (ref. 24 reviews early work), isotope exchange experiments, 7, 19 a new interpre-
tation of transpeptidation reactions, 20, 21 cryo-

emistry, 22 and an analysis of the dependence of re-
anction on pH 23 led to the establishment of the noncovalent general acid-base model of cataly-

sis 24, 25–28 at the expense of the covalent acyl-


---

Received July 4, 1990; revision accepted June 14, 1991.

Address reprint requests to Dr. Tom L. Blundell, Depart-
ment of Crystallography, Birbeck College, The University of 

Abbreviations: ED, electron density; translational, librational, and 

wax analysis; $R_m$, isotropic temperature factor; rms, root-

mean-square deviation. The nomenclature by Schachter and 

Bergs is used to label ligand and enzyme binding sites 

Schechter, L. Bergs, A. On the size of the active site in pro-


162, 1967.
medidate during catalysis. The same conclusion was reached from the X-ray analyses of fungal enzymes complexed with dipeptidase inhibitors.2,3,5,35 The mode of binding of the inhibitors allowed modeling of the transition state of the hydrolysis. These studies together with kinetic experiments played a major role in elucidation of some aspects of the aspartic proteinase mechanism and in the design of potent inhibitors (see ref 34 for a review).

However, the crystallographic data were not sufficient to rationalize all experimental evidence from kinetic studies. For example, the structural details of transpeptidation reactions,2,3,5,35 the reasons for the differences in hydrolysis of short and long substrates, where $k_{\text{cat}}$ increases sharply but $K_{m}$ remains constant,2,3,5 and the precise nature of multiple inhibitor binding26 are still not clear. A tentative interpretation of these processes involved conformational changes in the enzyme. The crystallographic structures known at the time gave no indications of a major conformational change in the enzyme that might take place during catalysis. The only exception was a movement of the "flag" (β-hairpin 73–82) that covers the active site cleft. The "flag" has to open to allow the access of the ligand into the active site cleft and occupied a slightly different position in the complex than in the unliganded state.35

In this paper, we extended the description of conformational changes associated with inhibitor binding to endoasparaginase37 and with activation of pepino- gen to pepsin.2 We show that a large part of the conformational variability in the aspartic proteinase family as a whole can be accounted for by a rigid body movement of the C-domain (100–328) relative to the rest of the molecule. We discuss the relative orientations of the two rigid bodies in terms of an interface structure and TLS analysis of thermal motions. We use 1C inhibitor complex with endoasparagine to describe the consequences of the rigid body movement for the shape of the active site cleft and for the conformation and position of the ligand in the cleft. Finally, we discuss implications of these observations on the conformational variability for the mechanism of aspartic proteinases.

**Materials and Methods**

**Structures**

Structures analyzed in this paper include 15 endoasparaginase complexes (Table 1) and eight forms of five different aspartic proteinases (Table 1b). Endoasparaginase–inhibitor complexes can be crystallized in two different unit cells (Fig. 4): one isomorphous to native endoasparaginase and one not.35

**Definition of Rigid Bodies**

A difference distance matrix is frequently used to represent conformational differences between two related molecules.37 It is defined as the absolute difference between the two Cα distance plots of the molecule compared. A difference distance matrix was used to define rigid bodies for native and inhibitor bound endoasparaginase. The same division was later justified for other comparisons by corresponding difference distance plots, least-squares superpositions, and inspection on a graphics terminal. Where different enzymes were compared the alignment used was obtained by the program COMPARE.42

**Definition of Rigid Body Movement**

Rigid body movement is best described in terms of a screw transformation. A translation along a and a rotation around a line in space are specified, such that one rigid body moves from its relative position in the first molecule to its relative position in the second molecule. The parameters of screw transformation were obtained from least-squares superpositions: the first one was the superposition of the
TABLE II. Aspartic Proteinases Used in This Analysis

<table>
<thead>
<tr>
<th>Protein</th>
<th>Code</th>
<th>Resolution (A)</th>
<th>R-factor</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endothiop胭thin</td>
<td>4APE*</td>
<td>2.1</td>
<td>0.18</td>
<td>1.2</td>
</tr>
<tr>
<td>Endothiapepsis-CP-69,789</td>
<td>1E27*</td>
<td>1.8</td>
<td>0.16</td>
<td>27</td>
</tr>
<tr>
<td>Peptidase</td>
<td>2APE</td>
<td>1.8</td>
<td>0.14</td>
<td>5.4</td>
</tr>
<tr>
<td>Rhodarmin</td>
<td>2APE*</td>
<td>1.8</td>
<td>0.14</td>
<td>5.6</td>
</tr>
<tr>
<td>Rhodarmin-reduced bond inhibitor</td>
<td>1APE</td>
<td>1.6</td>
<td>0.19</td>
<td>10</td>
</tr>
<tr>
<td>Peptidase</td>
<td>M207N</td>
<td>2.8</td>
<td>0.18</td>
<td>9</td>
</tr>
<tr>
<td>Peptidase</td>
<td>PO73N</td>
<td>1.7</td>
<td>0.21</td>
<td>13</td>
</tr>
<tr>
<td>Cell chymotrypsin</td>
<td>CH5L</td>
<td>2.2</td>
<td>0.20</td>
<td>10</td>
</tr>
</tbody>
</table>

*These structures were deposited to Brookhaven Protein Data Bank. The same code is used here. **Peptidase pepfinder in a monomeric unit cell (M207N) is essentially identical to hexagonal setup (1APE) that was deposited to Brookhaven Protein Data Bank with the code 2APE. **These structures appear in the April 1990 release of the Brookhaven Protein Data Bank. The Brookhaven codes for chymotrypsin and peptidase are 3CMS and IFESI, respectively.

reference rigid bodies and the second one was the superposition of the moving rigid bodies starting from the orientation obtained in the first step. The rotation matrix and the translation vector acquired in the second step were expressed in terms of the screw parameters. 8-10 Each least-squares superposition started with a list of equivalent Cα atoms from the alignment by COMPARER but that list was then iteratively reduced until all Cα atoms in the list were at a distance of < 3.6 Å from their equivalents in the other molecule. This ensured that local distortions in loop regions did not influence the estimate of the global rigid body movement.

TLS Analysis

In the most general harmonic model of thermal rigid body motion in molecules, the motion is described by T, L, and S tensors. 46 Kinematically, it is possible to express this general motion as a sum of six simple independent motions: three translations and three screw motions. 47 A simplifying assumption was made here that motion of a rigid body can be described by one screw oscillation only each that the displacement vector is

\[ u = A (\alpha + (\mathbf{a} - A \mathbf{a})\mathbf{r}) \]  

The position of each atom in the rigid body changes as parameter \( \alpha \) assumes values between \(-\mathbf{m} + \mathbf{m} + \mathbf{r} \) with probability \( r(k) \), where \( r(k) \) is any probability density function with mean equal to 0. Parameter \( t \) determines the pitch of the screw; it is the translation corresponding to 1 radian rotation. Column vector \( \mathbf{n} \) is a unit vector on the screw axis. Matrix \( A \) contains Cartesian coordinates of the moving point \((x, y, z)\):

\[ A = \begin{pmatrix} 0 & z & -y \\ -z & 0 & x \\ y & -x & 0 \end{pmatrix} \]

Matrix \( A \) is an equivalent matrix consisting of coordinates of any point on the screw axis. The screw motion described in Eq. (1) results in the following T, L, and S tensors

\[ T = a^{\mathbf{MN}} \]

\[ L = a^{\mathbf{MN}} \]

\[ S = a^{\mathbf{MN}} \]

where \( M = (I - A \mathbf{a}) \) and \( N = \mathbf{a} \mathbf{a}^T \). \( \sigma \) is standard deviation of \( p(k) \) and measures a magnitude of the motion. Note that the ability of a certain screw motion to account kinematically for given T, L, and S tensors does not prove that this screw motion is the actual trajectory of the rigid body in the crystal lattice.

The isostrophic temperature factor \( B_{\text{iso}} \) of any atom can be calculated from the T, L, and S tensors using the following relations:

\[ U = T + ALAT + AS + S^TAL \]

\[ B_{\text{iso}} = 8\pi^2\text{trace}(U) \]

The least-squares refinement program RESTRAIN was used to calculate T, L, and S tensors for the two rigid bodies of the endothiop胭thin-CP-69,789 complex. The RESTRAIN program was first modified to allow discontinuous rigid bodies consisting of any type of atoms. Main chain atoms only were included in the calculation of the tensor. The initial tensors for the start of refinement were calculated from Eqs. (2) using screw motion parameters as determined from the comparison of the native and CP-69,789 bound endothiop胭thin (Table III). Ten cycles of refining T, L, and S tensors simultaneously with positions of all atoms and isotropic temperature factors for atoms other than those in protein main chain resulted in the refined T, L, and S tensors. Finally, the overdetermined system of Eqs. (2) was solved by Newton–Raphson and linear least-squares methods to obtain a single screw motion best accounting for the refined T, L, and S tensors. This screw motion is described by a position of the screw in space, \( A \), orientation in space, \( \alpha \), pitch of the screw, \( \mathbf{a} \), and standard deviation of rotation, \( \sigma \). If we assume a Gaussian distribution for rotation, then rotational displacement from an equilibrium position will be within \( \sigma \) for
66.7% of time. The details of this new method for analysis of dynamic rigid body motion based on X-ray crystallography (ref. 46 and A. Sali and D.S. Moss, in preparation).

Principal Components Analysis of Rigid Body Orientations

Principal components analysis is a standard clus-
tering technique that projects a set of points from
a higher to a lower dimensional space in such a
way that the first axis in the lower dimensions accounts
for the largest possible fraction of the variance among
the original data points. The second axis accounts
for the largest possible fraction of the remaining
variance, etc. If it happens that only a small number
of these principal axes account for a large fraction of
the original variance then considerable simplifi-
cation in the representation of original relationship
between the points is possible by plotting them in
the lower dimensional space.

The relative orientation of two rigid bodies can be
described by six independent parameters (cf. Eq.
(1). These parameters specify a point in a six-di-
ensional space. However, because they depend on
an arbitrary choice of the reference structure, such a
simple description of a rigid body orientation is bi-
ased. Bias can be removed by using, in turn, each
structure as a reference when describing relative
orientation is one protein and then joining indi-
vidual parameter sets. In our analysis of a seven mem-
bered family, seven sets of six coordinates were
joined in a 12-dimensional vector specifying an un-
biased relative orientation of the two rigid bodies in
one protein. These seven 12-dimensional vectors can
then be subjected to a standard principal compo-
nents analysis to reduce the number of dimensions
still accounting for a large fraction of the original
variance. Before the eigenvalues/eigenvectors prob-
lem was solved, the vectors were normalized, so that
the mean and variance of the rotational and trans-
slational components of the seven vectors were 0 and
1 respectively.47

Miscellaneous

Mainchain hydrogen bonds were found by pro-
gram DSSP48 using an energy cutoff of ~1.5 kcal/
mo1. Hydrogen bonds involving side chains were de-
efined using a simple 2.5 A cutoff for the distance
between any protein donor and acceptor atoms. This
cutoff was relaxed to 3.6 A when water molecules
were included in the analysis. Hydrogenic contacts
were obtained as in ref. 40. Solvent contact areas
of residues were calculated by the algorithm of Rich-
mond and Richards.49 The interface between the two
rigid bodies in endothiapepsin was defined as a
union of residues obtained by three criteria: (1) Res-
ides with at least one atom closer than 4.5 A to
any atom from the other rigid body. (2) Residues that
lose more than 1/2 of their solvent contact area
when the two artificially separated rigid bodies form
the folded structure. (3) Residues that are hydrogen
bonded to water molecules to any residue in the other
rigid body.

RESULTS

Fold of the Aspartic Proteinases

A conventional starting point for the description of
the architecture of a protein structure is a division of
the fold into two structurally similar lobes.6 However,
when the interface between these two lobes involves a
well defined flat sheet rather than a packed sheet, we
may alternatively divide the molecule into three parts
(Fig. 1). In this description, there are two mainly
β-sheet domains—the N- and C-domains—that
provide the active site cleft and are also related by a
diagonal symmetry. These are picked on top of the
third part, a central motif involving the inter-
lamin β-sheet and a helix at each end. The C-domain
is more clearly separated from the rest of the molecule
than the N-domain (Fig. 1). Superposition of the
C1 backbone of seven forms of five different aspar-
tic proteinases establishes that by far the largest
differences across the family as well as between dif-
ferent forms of the same enzyme involve the C-do-
mains (Fig. 1).

The axial symmetry implies similarity be-
 tween the conformation and orientation of the N-
and C-domains. However, there are very important
differences between these two domains that may ex-
plain why the conformational changes described
here do not occur symmetrically with respect to the
aspartic proteinase fold as a whole. First, the ne-
ighboring edge strands of the central motif (a, -2, -7)
and the N-domain (a, 15, -20) are much closer than
the equivalent strands in the C-domain (α', 179, -184,
b', 190-199). The average distance between juxtaposed C1 α-carbons of residues 1-7 and 17-13 is
5.5 Å whereas the equivalent distance in the C-do-
main is 6.6 Å. This results in the gap between the
central motif and the C-domain that does not occur
in the N-domain. Second, the conformational differ-
ences between the two domains include the presence
of the short β3 helix (269-274) in the C-domain
that interfaces to the central motif via the a-helix
304-310. There is no equivalent of the helix 269-
374 in the N-domain. Additionally, while the N-do-
mains can be viewed as composed of two β-sheets L1
and S1, the sheet S2 of the C-domain has a definite
identity at residues 243-249 and 286-288. These two
strands are not hydrogen bonded to each other,
contrary to the equivalent strands 80-87 and
198-99 in the N-domain. Third, a major difference
between the two domains is the absence of the long
β-hairpin " flap" (171-82) and a-helix 108-114 in the
N-domain. When the ligand binds, the " flap" over-
" examines to S2, S1, and L2 pockets whereas the helical
contributes to binding the F, residue. The " flap" is
just a short turn (241-244) in the C-domain and the
The definition of rigid body in endothiapepin

A different distance plot (Fig. 2) shows the conformational changes in endothiapepin when the oligopeptide inhibitor CP 60,799 binds into its active site cleft. If one neglects local distortions such as those involving the "tag" this plot shows that there are only two parts of the structure that retain their internal conformation but change their spatial relationship when the inhibitor binds. Thus the two rigid bodies in endothiapepin with respect to the inhibitor binding are defined. The first rigid body comprises residues 2 to 129 and 303 to 356 and the second rigid body consists of residues 190 to 302. This division is in agreement with the tripartite description of the serine proteases fold: the first rigid body comprises both the central motif and the N-domain and the second right body corresponds to the C-domain.

Definition of the Rigid Body Movement in Endothiapepin

The description of the rigid body movement in terms of the screw motion (Fig. 2) shows that the conformational change involved a ~ rotation around a negligible translation along the screw axis that joins approximately through the active site center (Table III, first line). The rms value for the two Cα-dimensions is 1.47 Å when only the central motif and N-domain are used in the least-squares superposition (Table III). This is decreased to 0.06 Å (31% of the original value) when the two Cα-dimensions are superposed separately (Table III). This large reduction justifies description of the conformational transition as a rigid body movement as opposed to distortion within the C-domain.

The movement corresponds to changes in the relative positions of the Cα atoms of up to 2.5 Å. Changes in the active site cleft become appreciable (~1.5 Å) at the S1 pocket formed by the α-helix 109-114; the cleft opens up after the "tag"-p-hairpin, this helix undergoes the next largest decrease in thermal mobility of all regular secondary structure elements when the
Fig. 3. Geometry of the rigid body movement in endostatin-inhibitor complex. The movement of the second rigid body (C-domain) relative to the first (monomer 1 and inhibitor) is quantified in terms of the centre motion. Thick black lines: endostatin; thin line, the second rigid body of the complex between endostatin and inhibitor. The orientation of the inhibitor binds. The average isotropic temperature factor of the helix falls from 25 to 16 Å². There is a further local movement of this helix of ~0.8 Å in roughly the same direction as the global rigid body movement. This indicates that this could be the contact point on the enzyme used to trigger the conformational change by the ligand in the active site cleft. However, the reasons for a decrease in the average isotropic temperature factor and a positional change of this helix cannot be attributed unequivocally to the ligand binding because some of the helix residues are in contact with a neighbouring molecule in the unit cell of the inhibitor complex but not in that of the native enzyme.

Figure 4 shows the magnitude of rigid body changes for all 15 endostatin-inhibitor complexes. In all cases, the orientation of the active axis is the same as for the example described above. It is apparent that two forms of endostatin exist. The first one with a very small rotation and translation is very close to the native endostatin; the second one with ~4° rotation and very small translation of ~0.3 Å is the form discussed above. Several conclusions can be easily made. First, there is no core complex between the two rigid parts of the C-domain and the relative orientation of the rigid bodies, except that the domain inhibitors tend to associate with lower displacement from native conformation within each group of complexes. Second, there is a clear but not an exclusive association of the rigid body movement of the non-isomorphous cryo-electric unit cell. Third, the presence of the F2 and F3 regions of the helix seem to be a necessary but not a sufficient condition for the rigid body movement to occur, since the absence of the F2 and F3 regions is the only outstanding feature of the B020 inhibitor that crystallizes in the non-isomorphous unit cell and exhibits very small rigid body movement.

Comparison of Aspartic Proteases

Figure 1 shows that most of the structural similarity in the aspartic protease family involves the complex is obtained from the superposition of its first rigid body with the third rigid body of its native endostatin. The inhibitory complex is shown as a representative of the predominantly non-isomorphous group of inhibitory complexes that includes CIP 68-792, inhibitor (Fig. 4).

Fig. 4. Magnitude of the rigid body movement in 15 endostatin-inhibitor complexes. A rotation and a translation component are specified for all 15 inhibitor complexes. In addition, the type of cryo-electric unit cell is evaluated as an empty core for an isomorphous and a filled square for a non-isomorphous unit cell.

C-domain which corresponds to the second rigid body in endostatin-inhibitor binding. Therefore, the rigid body movement analysis including difference distance plots and asymmetry fitting was extended from the endostatin-inhibitor complexes to the whole family. A detailed description of some of the comparisons is presented in Figure 5 and Table III. Difference distance plots (data not shown) and a large reduction in the rms deviations between the second rigid bodies when the rigid body movement is allowed. Table III confirms that the rigid body movement is also a convenient way of accounting for a large part of the differences within the aspartic protease family in a whole.

Principal components analysis (Fig. 5) of rigid body orientations in the aspartic protease family separates mammalian numbers from fungal ones. Within the mammalian group, the synopen of papain is separated from pepsin and chymotrysin. Within
TABLE II. Rigid Body Movement in Aspartic Proteinases

<table>
<thead>
<tr>
<th>Protein</th>
<th>Before axis1</th>
<th>Magnitude2</th>
<th>Right body rms (Å)</th>
<th>rms ratio 2 after 3 before</th>
</tr>
</thead>
<tbody>
<tr>
<td>compared</td>
<td>Direction</td>
<td>Displacement (Å)</td>
<td>8°</td>
<td>1°</td>
</tr>
<tr>
<td>APE-APPE</td>
<td>-408, -375, -175</td>
<td>2.6, -0.2, 3.4</td>
<td>4.0</td>
<td>0.09</td>
</tr>
<tr>
<td>APE-APEP</td>
<td>-330, -340, -340</td>
<td>1.6, -1.8, -0.5</td>
<td>9.9</td>
<td>0.44</td>
</tr>
<tr>
<td>APE-PMN</td>
<td>-344, -179, -919</td>
<td>2.9, -0.6, -0.6</td>
<td>17.6</td>
<td>-0.14</td>
</tr>
<tr>
<td>APE-POGP</td>
<td>-445, -197, -773</td>
<td>9.9, -0.6, -0.7</td>
<td>17.6</td>
<td>-0.14</td>
</tr>
<tr>
<td>APE-APAP</td>
<td>-306, -191, -813</td>
<td>5.5, 2.5, 5.4</td>
<td>3.9</td>
<td>0.06</td>
</tr>
<tr>
<td>APE-PAP</td>
<td>-135, -125, -972</td>
<td>5.8, -3.8, -0.4</td>
<td>6.1</td>
<td>-0.03</td>
</tr>
<tr>
<td>MON-PON</td>
<td>-406, 383, 389</td>
<td>2.1, 0.4, -0.5</td>
<td>6.5</td>
<td>-0.45</td>
</tr>
<tr>
<td>MON-PON</td>
<td>125, 383, 704</td>
<td>-6.5, 0.1, -1.4</td>
<td>4.0</td>
<td>0.12</td>
</tr>
<tr>
<td>PON-CYN</td>
<td>305, -244, -121</td>
<td>0.3, 0.3, 0.3</td>
<td>5.7</td>
<td>0.10</td>
</tr>
</tbody>
</table>

*All screw axes are directly comparable: they were transformed applying the transformation from the stereoplot's of the first rigid body of the first molecule in the comparison to the endolysin's (4APE) first rigid body.

1Direction cosine and location of the axes are specified. The location is described by a displacement vector of the point on the axis closest to 0, where point 0 was found by minimizing the max of squared distances from it to all the axes. In the APE coordinate system, the axis 3 is at 0.5° away from 0, atoms 7 residues 213, 215, and 308.

2A retains its normal, and translation (t) along the screw axis is specified.

3C atoms fixed equivalent in the least 95% of carboxylated rigid bodies were used in both rms calculations for the second rigid body.

The rms deviations for the first rigid bodies as obtained from the least squares superposition of the first rigid bodies only. The second number in parentheses (N) are the numbers of equivalent C atoms in the number of initially equivalent residues as inferred from the alignment obtained by the program COMPARE20 and N is the final number of equivalence obtained from the refinement of the rigid body superpositions by iterative least-squares fitting (see also the section Definition of rigid body movement in the Methods).

The relative mean differences between the two number demonstrate that the larger parts of the enzymes do move as rigid bodies.

The rms deviation for the second rigid body as obtained from the superposition of the first rigid bodies only.

The rms deviation for the second rigid bodies as obtained from the superposition of the second rigid bodies only.

Fig. 5. Primal components analysis of the rigid body orientation in the Aspartic proteinase family. The two first principal components, a and b, acquire, by 94 and 4%, respectively, of the original variance among the seven orientations examined. For both principal components, loads of original rotation and translation coordinates are approximately the same; therefore, to simplify the representation, we have represented the coordinate by replacing some rotations or translations in Prime. Relationships described in Table III are indicated by a line connecting the two pegs in the comparison. Any axis of rotation that shows improvement is marked by an arrow. The second rigid body when rigid body movement is allowed are also shown—see Table III for further details.

the fungal group, endopeptidases forms are separated from rhinopapain and penicillopepsin, which are also significantly closer to mammalian enzymes than endopeptidase forms. This could contribute to the failure of endopeptidase and success of penicillopepsin and rhinopapain as the search modes for the molecular replacement technique in the X-ray analysis of Pepsin. 25° In the whole family, penicillopepsin and rhinopapain are the only enzymes with identical rigid body orientation, as judged by the small rotation angle and only small improvement in the rms distance when the movement is allowed. All other enzyme pairs have significantly different rigid body orientations. The magnitude of the rigid body rotation in the family ranges from 2° for penicillopepsin—rhinopapain to 17.6° for endopeptidase-pancreatin comparison. The latter is equivalent to 11 Å positional differences for C atoms distant from the axis. In all cases, translation is negligible compared to rotation. To put these numbers into perspective, a survey of the families of multidomain proteins showed that usually the differences in relative orientation of domains involve less than 1° rotation and improve the rms distance
by less than 30% (A. Sali, unpublished results; compare with Table III).

The complex between an oligopeptide inhibitor and rhinopepsin\(^2\) is the only aspartic proteinase-inhibitor complex besides endothiapepin complexes that was deposited to Brookhaven Protein DataBank. Rigid body movement in rhinopepsin (Table III), if any, is comparable in magnitude to the endotheiapin complexes with the native-like orientation of rigid bodies (Fig. 4). The only large conformational difference involves the "flag" region moving relative to the rest of the enzyme. The rigidity in rhinopepsin may be provided by the crystal packing: the rhinopepsin complex was prepared by soaking the inhibitor in the formed enzyme crystal whereas all endotheiapin complexes were co-crystallized.

Although the precise geometry and magnitude of the rigid body movement in pairwise comparisons of aspartic proteinases vary from case to case (Table III), the main features of the movement are the same. First, the hinge division in the two rigid bodies applies. Second, screw axes always axis rough enough through the interface between the rigid bodies in the vicinity of the two main chain segments that contact the two parts (Table III). These two parts also happen to be near the tips of the two loops carrying the essential aspartates 32 and 210. Third, the movement is associated with geometrical changes that are largest around the S3 pocket.

TLS Analysis

The T, L, and S tensors defined by RESTRAN\(^{24,25}\) for both rigid bodies in endothiapepin-\(^\text{TP-69,799}\) complex were obtained as described in the Methods. Figure 5 shows that the isotropic temperature factors calculated from these tensors (Eqs. (3)) have very similar values to those obtained directly from an isotropic temperature factor refinement.\(^{26}\) This indicates that a significant fraction of the protein dynamics and static disorder can be described by a rigid body motion.

The single screw motion calculated from Eqs. (3) accounts for the isotropic temperature factors of the enzyme (Eqs. (3)) almost as well as the original refined T, L, and S tensors (Fig. 6). This indicates that the rigid body motion can be at least kinematically approximated by a single screw oscillation.

The orientations and positions of the screw axes obtained from the TLS analysis for the two rigid bodies are very similar to those of the axis used to describe the static change for the first rigid body, the angle between the two screw axes is only 16° and the distance 1.2 Å (the standard deviation of rotation around the screw is 0.57, translation is negligible, the numbers for the second rigid body are 40° and 0.9 Å (the standard deviation for rotation is 2.8°). These results are consistent with the two rigid bodies of the molecule oscillating in a manner that leads most directly to the other equilibrium orientation (Figs. 3 and 4) when enough energy is accumulated to climb the activation energy barrier. The single screw-axis model derived from the refined T, L, and S tensors describes atomic displacements in the enzyme very well for each of the two rigid bodies. From Figure 6 it is clear that this very simple model accounts for 59% of the mean-square displacement observed with the full isotropic least-squares model which requires 2389 thermal parameters.

It is generally the case that the isotropic temperature factor in proteins correlates very strongly with solvent accessible volume or distance from the molecular gravity centre. The reason for this correlation is thermal flexibility of exposed smaller parts of the structure such as loops and not motion of a larger rigid body as a whole. TLS analysis cannot distinguish on its own between the two phenomena. As expected, the cross-correlation coefficient between the directly refined isotropic temperature fac-
The two rigid bodies of apartic proteins comprise separate parts of the structure and the boundary between them evident, both from the trace of C backbone (Fig. 1) and from the two-dimensional plot of hydrophobic contacts (Fig. 2). The interface between the rigid bodies in endothiopsin consists of 12 potentially charged residues (9 negatively, 3 positively), 15 polar, and 60 nonpolar residues (Fig. 7). As an example, these 80 residues represent a significant fraction of the total of 336 residues. The rigid bodies are held together by two main -chain segments of residues 180 and 310, by van der Waals contacts, and a side chain-side chain hydrogen bond, a main chain-mmain chain hydrogen bond, 10 side chain-main chain chrohydrogen bonds, and a number of hydrogen bonds mediated by 13 water molecules. There are no polar contacts at the cutout distance of 5 Å. Differences between the CP-D polypeptide and the native enzyme are limited to absence of three of the 13 bridging water molecules, to addition of seven new water molecules, and to minor changes in side chain of van der Waals contacts distances.

The most intense contacts is made between the tips of the two active site loops containing catalytic aspartates 32 and 215. The side loops are engaged in a side chain-mainchain hydrogen bonding network called a "firstman" grip. Additionally, there is a contact between the helix 303-309 from the first rigid body and the helix 265-274 from the second rigid body. While the disposition of the active site loops remains essentially unchanged during the movement as a result of the screw axis passing through them, a change is the relative orientation of the two helices that can be rationalized by the helix interface shear mechanism. The shear mechanism was proposed to be a major way of accomplishing the conformational variability in α-helical proteins by allowing differences of up to 1.5 Å in a relative helix orientation brought about by small adjustments between the side chains in contact. This description fits precisely the changes in the relative orientation of the two helices as well as the whole rigid bodies (Figs. 6 and 8). Additionally, the unfavorable atomic contacts resulting from the rigid body movement are dissipated further by very small movement of the 303-309 helix relative to the first rigid body (Fig. 2).

Consequences of the Rigid Body Movement for the Ligand Position and Conformation

Comparison of the positions and conformations of the inhibitors bound to different enzyme forms, each with a different orientation of the rigid bodies, provides an opportunity to study how the rigid body movement affects the relative position and conformation of ligands in the active site cleft.

To increase the accuracy of the comparison, the complexes were first superposed on their C-domains and the averages of positions for main chain atoms within the two groups of inhibitors were calculated (Fig. 9). For comparison, these averages were used instead of the individual structures. The averaging was justified a posteriori by the deviations of the main chain atoms from the average being in the order of the experimental error in coordinates smaller (≈ 0.25 Å).

Relatively large rms deviation of 0.73 Å for the two averages means that the positions and/or conformations of the two ligand averages relative to the C-domain are significantly different. The movement of the inhibitor main chain relative to the C-domain can be described as a "re«« motion involving a 0.34 Å translation along and a 4.7° rotation around the screw axis. When the movement relative to the larger rigid body is considered the 0.13 Å translation and 3.8° rotation occur. A relatively small rms deviation of 0.53 Å after the superposition of the inhibitor averages shows that the conformations of the two ligand types are not very different. However, there are real differences is supposed by the fact that this rms deviation is still linear in some expected by the average the rms error in the atomic positions of individual inhibitors is 0.45 Å; the then the expected rms deviation in the positions the five averages in 0.11 Å (0.55√5) and the expected rms deviation between two such averages is only 0.16 Å (0.5√5). The movement of the inhibitor as a whole and the internal distortion relative to real positional change of up to 1 Å for the main chain atoms of the P1 residue.

DISCUSSION

In the previous section, it was shown that a relatively large change in the shape of the active site cleft in endothiopsin occurs as a result of the rigid body movement of the C-domain relative to the rest of the molecule. It was also demonstrated that the
name type of a conformational change accounts for about 50% of the variability in three-dimensional structures among different enzymes in the aspartic proteinase family. Further, the TLS analysis of thermal motion in the endothiapipepin-CP-69,799 crys-

tal showed that atomic mobility is consistent with the division into two rigid bodies oscillating around the screw axis that are similar to the screw axis transforming one endothiapipepin form into the other. And finally, this rigid body movement may be
associated with changes in the conformation and relative position of up to 1 A of the inhibitor in the active site cleft. These differences may be even larger in a substrate that does undergo a change in the stereochemistry of the scissile peptide bond.

In endostatipain and pepstatin, the conformational change is associated with the presence of the scissile peptide in the active site cleft. Inhibitors in endostatipain and pepstatin, although other factors such as crystalline packing forces may also be responsible. Unfortunately, it is not yet possible to choose between the two alternatives for the driving force of the conformational transition: ligand binding or crystal environment. The crystallization of the BW258 complex with endostatipain in the noncrystalline unit cell, yet with the native-like orientation of rigid bodies, indicates that it is the absence of the SB and SP residues that is the only outstanding feature of the BW258 inhibitor. The data show that small conformational changes similar to those occurring in the rigid body movement occur in all endostatipain complexes (Fig. 4) and that the noncrystalline and noncrystalline pepstatin forms have the same conformation (A. Salt, unpublished) also point to the local binding as a likely factor contributing to the conformational change. On the other hand, the crystallization of the 18181 complex in the crystalline and noncrystalline forms with native and nonnative-like orientation of the rigid body, respectively, indicates that the conformational change is either a sequence or an accompanying feature of a cortical crystal environment.

Even if crystallization is the relative orientation of the two rigid bodies, the fact remains that the fold of endostatipain and probably of the other aspartic proteinases can exist in a delicate equilibrium between two forms differing considerably in the shape of the active site cleft. This is also a fact that the two forms are able to interconvert as a result of a stimulus from the environment, possibly as a result of filling in the SB pocket. In this context, it is reasonable to propose a working hypothesis that this structural flexibility plays a part in the function of the aspartic proteinases. It could serve to increase both the turnover and specificity in the hydrolysis of peptide bonds.

It has been suggested before that some of the steps in the catalysis by aspartic proteinases involve a conformational change in the enzyme. This was based mainly on the 5000-fold increase in $K_M$ and virtually no change in $K_a$ when the SP residue is added to the substrate. However, the same effect could also be due to nonproductive binding of short substrates or distortion of the substrate itself.

The enzyme distortion hypothesis was that the $K_M$ residue by occupying the SB pocket, uses more of the binding energy to trigger the conformational change in the enzyme, a change which decreases the activation energy of the rate-determining step and thereby increases $K_a$. There are several pieces of evidence which support this view. The Arrhenius plots of the hydrolysis of few substrates have a sharp break which has been shown to correlate with a break in the temperature dependence of the intensity of a circu- lation dichroism band at 342 nm in a pepstatin complex. The Arrhenius plots in linear for short substrates, as is the temperature dependence of the 342 nm circular dichroism band of the free enzyme. The break in the Arrhenius plot is thus most likely due to a conformational change in the enzyme. The 342 nm band is an unusual band which is common to all three fungal aspartic proteinases as well as pepstatin and originates from a nitrilase feature that is very sensitive to raised temperatures.
pH and urea concentration (T. Hofmann, unpublished results). It is therefore associated with parts of the molecule that are fairly flexible. The intensity of this band also changes upon binding of peptin. The following earlier observations are also consistent with a role of enamine conformational changes in the function of asparagine proteases. First, a multistep inhibitor binding occurs when the P2 residue is present while the kinetics are simpler for inhibitors without a P2 residue. Second, pre-steady-state fluorimetric measurements indicate conformational changes. Third, small nonpeptide substrates can act as activators of the hydrolysis of poor short substrates, probably by simulating the binding in P3 pocket, as an event that is accompanied by the changes in the CD spectrum which are probably due to conformational changes.

We propose that at least some of these solution experiments may be rationalized with the help of the crystallographically observed rigid body movement described in this paper. Apart from the role that the rigid body movement in aspartic proteases may play in the hydrolysis, there are two additional potential applications. The first one is in the design of potent inhibitors: an inhibitor that did not induce a conformational change in the enzyme would probably bind better, since it would not impart any driving energy on distorting the enzyme. The second application is in modeling of the structures of as yet unknown aspartic proteases and their zymogens; when a rigid body movement is taken into account considerably better models are obtained. These models can be used, for example, in a molecular replacement technique of X-ray crystallography.

NOTE


ACKNOWLEDGMENTS

We are indebted to Dr. Matthew Newman and Dr. Jim Bevington and Jan Hartscz for providing us with the structure of chymotrypsin and peptinogen, respectively, prior to their publication. A.S. was funded by GRR X-ray scheme, Research Council of Slovenia, Merck Sharp & Dohme and J. Stefan Institute, Ljubljana, R.V. was funded by the SERC and J.B.C. of the MRC.

REFERENCES

