Normal-mode analysis suggests important flexibility between the two N-terminal domains of CD4 and supports the hypothesis of a conformational change in CD4 upon HIV binding

Y.H. Sanejouand
Laboratoire de Physique Quantique, UMR 5626 of CNRS, IRSAMC,
Université Paul-Sabatier, 118 Route de Narbonne, 31062 Toulouse Cedex, France

Human CD4 is the receptor for human immunodeficiency virus (HIV). It is well established that the first domain of CD4 binds with high affinity to gp120, an envelope protein of HIV, but it has also been demonstrated that amino acids located in its second domain, within or close to residues 120–127 or 163–166 (lying 15 Å away from the binding site), play a role in virus infectivity. We show here that these two stretches of amino acids happen to be important for the largest amplitude motion obtained with the normal-mode theory for the two N-terminal domains of human CD4: an overall rigid-body displacement of one domain with respect to the other. Such a ‘hinge-bending’ motion is unexpected since these two domains were found by crystallographers to be tightly abutting. On the other hand, since for several proteins the hinge-bending motion experimentally observed upon ligand binding was found to be similar to the largest amplitude motion obtained with the normal-mode theory for these proteins, our results suggest that CD4 may undergo such a kind of conformational change upon HIV binding.

Keywords: gp120/HIV-blocking monoclonal antibodies/human immunodeficiency virus/second domain of CD4/virus entry

Introduction

Gp120 binding to the N-terminal domain of human CD4 is the first step of a pH-independent process leading to HIV entry into cells and to its subsequent replication (Stein et al., 1987; Maddon et al., 1988; McClure et al., 1988). Membrane-expressed gp120 and CD4 interaction also mediates syncitia formation, a process of cell fusion leading to in vitro destruction of HIV-infected cells (Lifson et al., 1986; Sodroski et al., 1986). Both processes, still poorly understood, are considered to be complex. For instance, in human cells, surface expression of CD4 is necessary, and sufficient, for viral infection in vitro. However, in murine cells, expression of CD4 from human origin allows virus binding but not infection (Madden et al., 1986).

It has often been proposed that one of the steps following gp120–CD4 interaction might be a conformational change in the HIV envelope glycoproteins, induced by gp120 binding to CD4 (Kowalski et al., 1987; Allan et al., 1990; Allan 1991; Moore et al., 1992). Evidence for an HIV-1-induced conformational change in CD4 has also been obtained. For instance, a monoclonal antibody raised in a mouse by human CD4 immunization was found to exhibit a greater affinity for human CD4–gp120 complexes than for human CD4 alone (Healey et al., 1990). Other such antibodies were obtained by immunizing mice with CD4–gp120 complexes (Celada et al., 1990; Gershoni et al., 1993) and, recently, the Fab 3–47 monoclonal antibody was found to recognize cell surface-expressed CD4 following the incubation of these cells with gp120 or with HIV-1 virions, whereas it does not bind to CD4+ cells, so gp120 or to cell surface-expressed HIV-1 envelope glycoproteins (Bachelder et al., 1995).

The purpose of the present work is to give a picture, qualitative but at the residue level, of the conformational change occurring in CD4 upon gp120 binding. One way to obtain a precise picture would be to crystallize a gp120–CD4 complex, determine its three-dimensional structure using X-ray crystallographic methods and compare it with the corresponding structures of free gp120 and CD4. For human CD4, the structure of the two N-terminal domains (residues 1–178) has been solved (Ryu et al., 1990, 1994; Wang et al., 1990), but attempts to obtain the structures of both the gp120–CD4 complex and free gp120 have failed so far. Here, the main tool used is normal-mode analysis, one of the best suited theoretical methods for studying collective motions in proteins. This theory has recently been shown to give a better picture than molecular dynamics methods of X-ray diffusive scattering of protein crystals (Faure et al., 1994), despite the fact that it is based on a severe small displacements approximation (see below). However, the idea that normal-mode theory is an accurate tool for the purpose of the present study comes from the fact that in many cases, e.g. hexokinase (Harrison, 1984), hen egg white lysozyme (Brooks and Karplus, 1985), human lysozyme (Gibrat and Go, 1990), myoglobin (Seno and Go, 1990a,b) and citrate synthase (Marques and Sanejouand, 1995), the largest-amplitude motion obtained with this theory was found to compare well with the conformational change observed by crystallographers in these proteins upon ligand binding (McCammon et al., 1976; Bennett and Stetitz, 1980; Remington et al., 1982; Huber and Bennett, 1983).

Hereafter, following a summary of the methodology, the largest amplitude motion of the two N-terminal domains of human CD4 obtained with the normal-mode theory is analysed and interpreted in detail. Then, experimental data supporting the hypothesis that CD4 actually undergoes, upon gp120 binding, a conformational change similar to this motion are examined.

Methodology

The normal-mode theory is based on the following ideas (Goldstein, 1950). In the vicinity of a stationary point, the potential energy of a system, \( V \), can be approximated by

\[
V = \frac{1}{2} \sum_{i,j} k_{ij}(r_i - r'_i)(r_j - r'_j)
\]

where \( k_{ij} \) are the second derivatives of the potential energy with respect to coordinates \( r_i \) and \( r_j \) and where \( r'_i \) and \( r'_j \) are the \( i \) and \( j \) coordinates of the stationary structure. Within this approximation, the equations of motion for a system of \( N \) atoms can be solved analytically, leading to the following solutions:
\[ r_i(t) = r_i^0 + \frac{1}{\sqrt{m_i}} \sum_{j=1}^{3N} C_j \cos(\omega_j t + \Phi_j), \quad i = 1, 3N \]  

which means that each atomic motion results from the superimposition of 3N independent sinusoidal contributions—the normal modes. In these equations, m_i is the mass of atom i, \( \nu_j = \omega_j / 2\pi \) the frequency of normal mode j, \( \omega_j \) the jth eigenvalue of the 3N×3N mass-weighted second derivatives of the potential energy matrix and \( \mathbf{a}_j = (a_{j1}, a_{j2}, ..., a_{jn}) \) the jth eigenvector of this matrix. \( C_j \), the amplitude of normal mode j, is such that

\[ C_j = \frac{\sqrt{2E_j}}{2\nu_j} \]

where \( E_j \) is the amount of energy in mode j. Applying the equipartition principle leads to \( E_j = k_b T \) (\( k_b \) being the Boltzmann constant and T the absolute temperature), which implies

\[ C_j = \frac{\sqrt{2k_b T}}{2\nu_j} \]

One can see that, at a given temperature, the lower the frequency of a normal mode, the larger its amplitude. Typically, for proteins, the normal modes whose frequencies lie under 30–100 cm\(^{-1}\) are found to be responsible for most of the amplitude of the atomic displacements (Levy et al., 1982; Swaminathan et al., 1982).

Starting from 3cd4, the better resolved (\( R = 2.2 \) Å) X-ray structure currently available in the Brookhaven Protein Data Bank (Bernstein et al., 1977), the potential energy was minimized with the CHARMM 21.3 program (Brooks et al., 1983), using ‘extended atoms’ and the same standard parameters and options as in previous methodological studies (Durand et al., 1994; Marques and Sanejouand, 1995). In particular, a cutoff at 7.5 Å was used in the calculation of non-bonded interactions, with a shifting function for electrostatics, and a switching function between 6.5 and 7.5 Å for Van der Waals interactions (Brooks et al., 1983). The minimization process was stopped at a gradient root-mean-square value of 0.0001 kcal/mol Å, the \( C_a \) root-mean-square deviation from the X-ray structure being 1.9 Å. Then, in order to compute the lowest frequency normal mode of CD4, i.e. the largest-amplitude motion predicted by the normal-mode theory for this protein, the matrix of the mass-weighted second derivatives of the potential energy was diagonalized. Since in the present case this matrix is not very large (3N = 5166), standard methods were used, on the IBM-SP2 facility available in CNUSTC (Montpellier, France).

Results and discussion

Normal-mode calculation

The three lowest frequency motions (1.1, 2.1 and 3.0 cm\(^{-1}\)) of the two N-terminal domains of human CD4 are found to be dominant hinge-bending motions; this is well shown by comparing, at a given temperature, overall atomic displacements with intra-domain atomic displacements for each low-frequency normal mode motion (see Figure 1). However, as expected from Equation 2, the amplitude of the 1.1 cm\(^{-1}\) motion is much larger.

The frequencies of the normal modes previously found to compare well with experimentally observed hinge-bending motions of proteins upon ligand binding happen to be signific-

![Fig. 1. Maximum \( C_a \) root-mean-square (r.m.s.) displacements when 2 kcal/mol are given to a single normal-mode motion, as a function of its frequency. Displacements are computed according to Equation (1), the amplitude of the considered normal mode being given by Equation (2), while the amplitudes of the others are set to zero. Full line: overall r.m.s. displacements. Dotted and thin lines refer, respectively, to intra-domain displacements in domains 1 and 2, that is, to r.m.s. displacements when only \( C_a \) displacements in one domain are considered and when the orientation of this domain is kept constant. The difference between both kinds of r.m.s. displacements is due to the rigid-body displacement of one domain with respect to the other, that is, to a ‘hinge-bending’ motion.](image)
several N-terminal residues (Val3, Leu5, Gly6, Gly9, Glu13, Leu14) as well as for some residues in loop BC of domain 2 (Pro122, Gly123, Ser125). On the other hand, the solvent accessibility of several other residues varies by a large amount during the hinge-bending motion within or near loop FG of domain 2 (Leu162, Gln163) (see Figure 3c).

Finally, in Figure 4, a residue-residue distance variation map is displayed. In such maps (Seno and Go, 1990a), white areas near the diagonal indicate which blocks of amino acids behave as rigid bodies during the motion. Here, the fact that the two domains of CD4 are moving relative to each other is made particularly clear, in addition to the movements of loops BC and FG relative to the rest of domain 2 during the hinge-bending motion.

Experimental evidence for a role of the hinge-bending motion in CD4 function

The hypothesis that a conformational change similar to the motion described above is important for the recognition process of Gp120 by CD4 and for the subsequent entry of HIV in lymphocytes is consistent with a fairly large body of experimental data, as summarized in Figure 5. First, the only mutations in domain 2 which were found to lower the affinity of CD4 for gp120 are located either in loop BC [Pro121, Pro122, Gly123 → Ser, Lys, Val (Clayton et al., 1988)] or in loop FG [insertion of Ser–Arg after Asn164 (Mizukami et al., 1988)]. Interestingly, when the Gp120 binding site is constructed in rat CD4 by replacing in its sequence residues 33–58 with residues taken from human CD4 sequence, the affinity of this chimera for Gp120 is found to be twofold less than that of human CD4 (Schockmel et al., 1992).

Second, residues in loops BC and FG have been shown to lie within or close to the epitopes recognized by several HIV-blocking monoclonal antibodies (Sattentau et al., 1986; Jameson et al., 1988; Mizukami et al., 1988; Peterson and Seed, 1988; Ashkenazi et al., 1990; Rubel et al., 1992; Moore et al., 1992), as detailed in Table I. For the well studied MT151 anti-CD4 monoclonal antibody (Sattentau et al., 1986; Jameson et al., 1988; Mizukami et al., 1988; Peterson and Seed, 1988; Ashkenazi et al., 1990), the residues found to belong to the epitope it recognized are either in domain 1 or in domain 2, on both sides of the hinge axis (Lys1, Gln94, Gln165 in loop FG). The binding of MT151 to CD4 should
that the HIV-blocking activity of 5A8 is also achieved by blocking a conformational change in CD4 similar to the hinge-bending motion described in Figures 2–4.

Such a conformational change could also be involved in the normal function of CD4 since the residues of domain 2 whose mutations were shown to affect CD4–MHC class II interactions (Leahy, 1995) are Gln165 (Fleury et al., 1991), Lys166, Lys167 and Lys171 (Moebius et al., 1991, 1992). These residues happen to be located within or close to loop FG.

Other evidence for inter-domain flexibility in CD4

Actually, a 1.3° amplitude hinge-bending motion of CD4 has already been observed when two X-ray structures obtained in different crystal packing environments were compared (Ryu et al., 1994). This was also observed in the course of a 165 ps molecular dynamics simulation in water, which exhibits a negative correlation between the atomic motions in the two domains (Ptaszek et al., 1994). Because a tight association between the two domains was observed in all structural studies (Ryu et al., 1990, 1994; Wang et al., 1990), the occurrence of a large amplitude motion had been discarded. One contribution of the present work is to show that important inter-domain flexibility is possible in spite of such a tight association.

Actually, several residues lying in the hydrophobic interface between the two domains (Ryu et al., 1990), are now found to be important for the hinge-bending motion (residues Val3, Leu5, Leu96, Pro121, Gln163). This suggests that the flexibility of some of these residues is fairly high. Such is certainly the case with residues in loop BC, in which there is, together with one glycine, a high proportion (3/8) of proline residues—a feature frequently observed in flexible regions of proteins (Braden and Toose, 1991).

Interestingly, loop BC is perfectly conserved among primates, i.e. among the species whose CD4+ cells were reported to be susceptible, in vitro, to HIV infection (McClure et al., 1987) (see Table II). Such a feature is also observed in the hinge region (residues 95–99; data not shown) and in the N-terminal region found to be flexible in the present study (residues 3–14; data not shown), but not in loop FG and around. There, only two or three amino acids over six are identical in primate species and in human or chimpanzee (cf. Table II). However, it is known from studies of chimeras of rat and human CD4 that rat CD4 with residues 33–62 taken from human sequence mediates HIV-1 infection (Davis et al., 1993). From the point of view of the present study, this suggests that a hinge-bending motion also occurs in these chimeras upon gp120 binding. Indeed, there is no reason to believe that CD4 function is different in primate and in non-primate species, namely, that large amplitude conformational changes may occur in some of these CD4 and not in others. On the other hand, in loops BC and FG, only Pro126 and Gln165 are found in all known sequences of CD4. Taken together with the results of the present study, this suggests that some mutations of these residues in human CD4 may have consequences on the susceptibility to HIV infection of lymphocytes bearing such modified CD4.

Second lowest frequency motion

The lowest frequency motion given by the normal-mode theory is not always the one reflecting the best rigid-body-like motions observed by crystallographers. For instance, it is the second lowest frequency motion of the T form of haemoglobin, which was found to be very similar to the observed difference between the T and R forms (Perahia and Mouawad, 1995).
Table I. CD4 epitopes recognized by HIV-blocking monoclonal antibodies MT151, OKT4B, OKT4F and 5A8, these epitopes being mapped either by escape mutations (Peterson and Seed, 1988; Ashkenazi et al., 1990), Ser–Arg insertions (Mizukami et al., 1988) or peptide competition experiments (Jainson et al., 1998; Burkly et al., 1992)

<table>
<thead>
<tr>
<th>Monoclonal antibody</th>
<th>MT151</th>
<th>OKT4B</th>
<th>OKT4F</th>
<th>5A8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epitope residues</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>94</td>
<td></td>
<td></td>
<td>99–105</td>
</tr>
<tr>
<td></td>
<td>165</td>
<td></td>
<td>165</td>
<td>121–124</td>
</tr>
<tr>
<td></td>
<td>139–168</td>
<td></td>
<td>139–168</td>
<td>127–134</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>165</td>
</tr>
<tr>
<td>Inhibition of HIV-induced cell–cell fusion</td>
<td>Strong</td>
<td>Weak</td>
<td>Strong</td>
<td>Strong</td>
</tr>
</tbody>
</table>

*Top, first domain residues; bottom, second domain residues.

Table II. Known sequences of loops BC and FG in domain two of CD4

<table>
<thead>
<tr>
<th>Species</th>
<th>Accession code in OWL database</th>
<th>Loop BC</th>
<th>Loop FG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primate species</td>
<td></td>
<td>120</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td>Human</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CD4_HUMAN</td>
<td>S.PPGSSPS</td>
<td>LQNQKK</td>
</tr>
<tr>
<td></td>
<td>Chimpanzee</td>
<td>S.PPGSSPS</td>
<td>LQDNKK</td>
</tr>
<tr>
<td></td>
<td>Red gibbon</td>
<td>S.PPGSSPS</td>
<td>SQDNKT</td>
</tr>
<tr>
<td></td>
<td>Green monkey</td>
<td>S.PPGSSPS</td>
<td>SQDNKT</td>
</tr>
<tr>
<td></td>
<td>Rhesus macaque</td>
<td>S.PPGSSPS</td>
<td>SQDNKT</td>
</tr>
<tr>
<td></td>
<td>Pig-tailed macaque</td>
<td>S.PPGSSPS</td>
<td>SQDNKT</td>
</tr>
<tr>
<td></td>
<td>Sooty mangabey</td>
<td>S.PPGSSPS</td>
<td>SQDNKT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>162</td>
</tr>
<tr>
<td></td>
<td>Rabbit</td>
<td>A46254</td>
<td>G.PSVGSPS</td>
</tr>
<tr>
<td></td>
<td>Dog</td>
<td>CD4_CANFA</td>
<td>N.PGSSPS</td>
</tr>
<tr>
<td></td>
<td>Mouse</td>
<td>CD4_MOUSE</td>
<td>SNSKSYMPL</td>
</tr>
<tr>
<td></td>
<td>Rat</td>
<td>CD4_RAT</td>
<td>SNPVYDPP</td>
</tr>
</tbody>
</table>

*The residue numbering comes from the human sequence, as found in the Brookhaven Data Bank (Bernstein et al., 1977). The sequences were taken from the OWL database, release 23.0, and a multiple alignment was performed with the GCG program package (Genetic Computer Group, 1994). From the structural point of view, loop FG is slightly shorter (see text).

For one monomer of dimeric citrate synthase, the lowest frequency motion of the closed form is similar to the observed difference between the open and closed forms but, when the other monomer is considered, it is the second lowest frequency mode which is the more similar to the observed difference (Marques and Sanejouand, 1995). In the later case, the conformational change observed by crystallographers is therefore much better reflected by a combination of the two lowest frequency motions. Since this might happen to be a general result, the analysis performed for the 1.1 cm⁻¹ motion of CD4 was also done for the 2.2 cm⁻¹ motion (Figure 6). Interestingly, the characteristics of this other type of hinge-bending motion (see Figure 1) are not much different from those of the 1.1 cm⁻¹ motion. In particular, residues in loop BC are also found to be among the most flexible (Ser120, Gly123, Ser124). Here, the hinge is located at Gly99, that is, exactly where domain 1 ends (Ryu et al., 1990; Wang et al., 1990). For Val97, ΔΦ and Δψ are large, but of opposite sign, which means that its motion is mainly a relative displacement of the side chain with respect to the backbone. The main differences between the 1.1 and 2.0 cm⁻¹ motions lie in the N-terminal region, more flexible in the former case, and in loop FG, whose accessibility to the solvent does not vary much in the latter case.

Other pairs of immunoglobulin-like domains

The structure of the two domains of the extracellular region of CD2 is very similar to the corresponding pair of domains of CD4, but they are less intimately associated than those of CD4—they are separated by a flexible linker region (Jones et al., 1992). Indeed, in the structures of rat and human CD2, the relative orientation of the two domains differs by up to 20° (Bodian et al., 1994). Note that in the structure of domains 3 and 4 of rat CD4, there is a 25° difference in their orientation when compared with domains 1 and 2 of human CD4, while 40% of the residues in the two pairs of domains are homologous (Brady et al., 1993; Lange et al., 1994). At a more general level, it is well known that immunoglobulin and the structurally similar fibronectin type III domains may adopt strikingly different relative orientations in a pair, despite the similarity of the individual domains (Campbell and Spitzfaden, 1994). The present study suggests that CD4 may adopt two (or several) such different conformations.

Conclusion

The normal-mode theory predicts that a large amplitude rigid-body-like motion of the two N-terminal domains of human CD4 may occur. One important stretch of amino acids found to be important for this hinge-bending motion is loop BC, in the second domain (residues 120–127), whose flexibility is fairly high during the hinge-bending motion. Others are the N-terminal region, the hinge region (near residue 99) and loop FG, also in the second domain (residues 163–166).

Some of the residues of loops BC and FG were previously found to be important for the high affinity of CD4 for gp120 [residues Pro121, Pro122, Gly123 (Clayton et al., 1988) and Asn164 (Mizukami et al., 1988)]. These, and others, were also shown to belong to the epitopes recognized by monoclonal
antibodies MT151, OKT4B, OKT4F and 5A8 (Sattentau et al., 1986; Jameson et al., 1988; Mizukami et al., 1988; Peterson and Seed, 1988; Ashkenazi et al., 1990; Moore et al., 1992). These antibodies, although they do not compete with gp120 for CD4 binding (their binding sites are far apart), block HIV-induced syncitia formation and/or HIV infectivity. The results of the present study suggest a way to explain how these antibodies achieve their HIV-blocking activity, namely by interfering with the hinge-bending motion of CD4. On the other hand, they strongly suggest that seeking molecules able to block this motion could become a powerful way for designing new anti-HIV agents.

Acknowledgements

The author thanks Simon Mathieu for his help with the matrix calculations. Georges Trinquier for fruitful discussions and the referees for their useful comments. Grants of computer time from the CNRSEC in Montpellier and from the Centre de Bioinformatique in Villejuif are acknowledged.

References


Received December 20, 1995; revised April 10, 1996; accepted April 18, 1996