A 3D Structure Model of Integrin $\alpha4\beta1$ Complex: I. Construction of a Homology Model of $\beta1$ and Ligand Binding Analysis

Tony J. You, David S. Maxwell, Timothy P. Kogan,* Qi Chen,† Jian Li, Jamal Kassir, George W. Holland, and Richard A. F. Dixon

Department of Medicinal Chemistry and Biophysics, Texas Biotechnology Corporation, Houston, Texas 77030 USA

ABSTRACT It is well established that integrin $\alpha4\beta1$ binds to the vascular cell adhesion molecule (VCAM) and fibronectin and plays an important role in signal transduction. Blocking the binding of VCAM to $\alpha4\beta1$ is thought to be a way of controlling a number of disease processes. To better understand how various inhibitors might block the interaction of VCAM and fibronectin with $\alpha4\beta1$, we began constructing a structure model for the integrin $\alpha4\beta1$ complex. As the first step, we have built a homology model of the $\beta1$ subunit based on the I domain of the integrin CD11B subunit. The model, including a bound Mg^{2+} ion, was optimized through a specially designed relaxation scheme involving restrained minimization and dynamics steps. The native ligand VCAM and two highly active small molecules (TBC772 and TBC3486) shown to inhibit binding of CS-1 and VCAM to $\alpha4\beta1$ were docked into the active site of the refined model. Results from the binding analysis fit well with a pharmacophore model that was independently derived from active analog studies. A critical examination of residues in the binding site and analysis of docked ligands that are both potent and selective led to the proposal of a mechanism for $\beta1/\beta7$ ligand binding selectivity.

INTRODUCTION

Integrins are cell-adhesion molecules expressed on the surface of cells, and they are involved in cell–cell and cell–matrix interactions (Adams and Lobb, 1999; Horwitz, 1997). Each integrin is a heterodimer composed of one α and one β subunit, and >20 $\alpha\beta$ pairs have been discovered (Chothia and Jones, 1997; Haas and Plow, 1994). The integrin very-late antigen-4 (VLA-4) is one of two known $\alpha4$ integrins, and it is the counter-receptor for the vascular cell adhesion molecule (VCAM-1) and the CS-1 splice variant of fibronectin (Elices et al., 1990; Wayner et al., 1989). Binding of VCAM-1 to $\alpha4\beta1$ is important in the inflammatory response, and blocking this interaction would be expected to be important in controlling many diseases (Irie et al., 1995). Understanding the nature of the interaction between the integrin and cell adhesion molecules can aid in the design of potent therapeutics. It has been shown that small cyclic peptides based on segments from either CS-1 or VCAM-1 can inhibit the VCAM-1–integrin interaction (Vanderslice et al., 1997; Wang et al., 1995), and these peptides were used to design very potent small molecular therapeutics (Adams and Lobb, 1999; Jackson et al., 1997; Kc et al., 1999). Antibody binding studies and mutagenesis data also point to the involvement of both the α (Schiffer et al., 1995; Loftus et al., 1994; Kamata et al., 1995; Irie et al., 1995) and β (Kamata et al., 1995; Puzon-McLaughlin and Takada, 1996) subunits in binding. However, the detailed mechanism of ligand–receptor interactions remains unclear due to the lack of a three-dimensional (3D) structure of the $\alpha4\beta1$ complex.

In some integrin α subunits, there is a special residue sequence referred to as the I domain, and there is growing evidence for its importance in ligand binding (Loftus et al., 1994; Lee et al., 1995a). The I domain is a sequence of $\sim$190 residues near the N terminus of the α subunit that is homologous to the A domains of von Willebrand factor (VWFA) (Lee et al., 1995a). The I domains known to date adopt a dinucleotide-binding fold and contain a metal ion-dependent adhesion (MIDAS) motif with a bound dication (usually Mg^{2+}) at the top of the β sheet (Lee et al., 1995b; Emsley et al., 1997). A number of groups have proposed that a topology similar to the I domain is present within the integrin β subunit (Puzon-McLaughlin and Takada, 1996; Tuckwell and Humphries, 1997). Based on detailed secondary structural comparisons, Tuckwell and Humphries concluded that the ligand-binding region of integrin β subunits is homologous to the N terminal region of the VWFA domain. Here, we describe the work of constructing a $\beta1$ homology model based on the I domain of integrin CD11B/CD18 with bound Mg^{2+} (PDB entry 1ido) (Lee et al., 1995b). This was the first step of our effort to build a complete $\alpha4\beta1$ complex structure model aimed at exploring the landscape of the interactions between the cell adhesion molecules and their inhibitors. The refined $\beta1$ homology model was used to study the binding mechanism of some de novo synthesized compounds and to verify a pharmacophore model previously derived from active analog analysis (Holland, 1999). Based on the ligand binding study, a possible mechanism behind ligand selectivity between $\alpha4\beta1$ and $\alpha4\beta7$ is proposed.

Received for publication 29 May 2001 and in final form 20 August 2001.

*Deceased.

†Current address: Lilly Research Labs, Eli Lilly and Company, Indianapolis, IN 46285.

Address reprint requests to Tony You, Texas Biotechnology Corp. 7000 Fannin, Suite 1920, Houston, TX 77030. Tel.: 713-796-8822 ext 132; Fax: 713-796-8232; E-mail: tyou@tbc.com.

0006-3495/02/01/447/11 $2.00

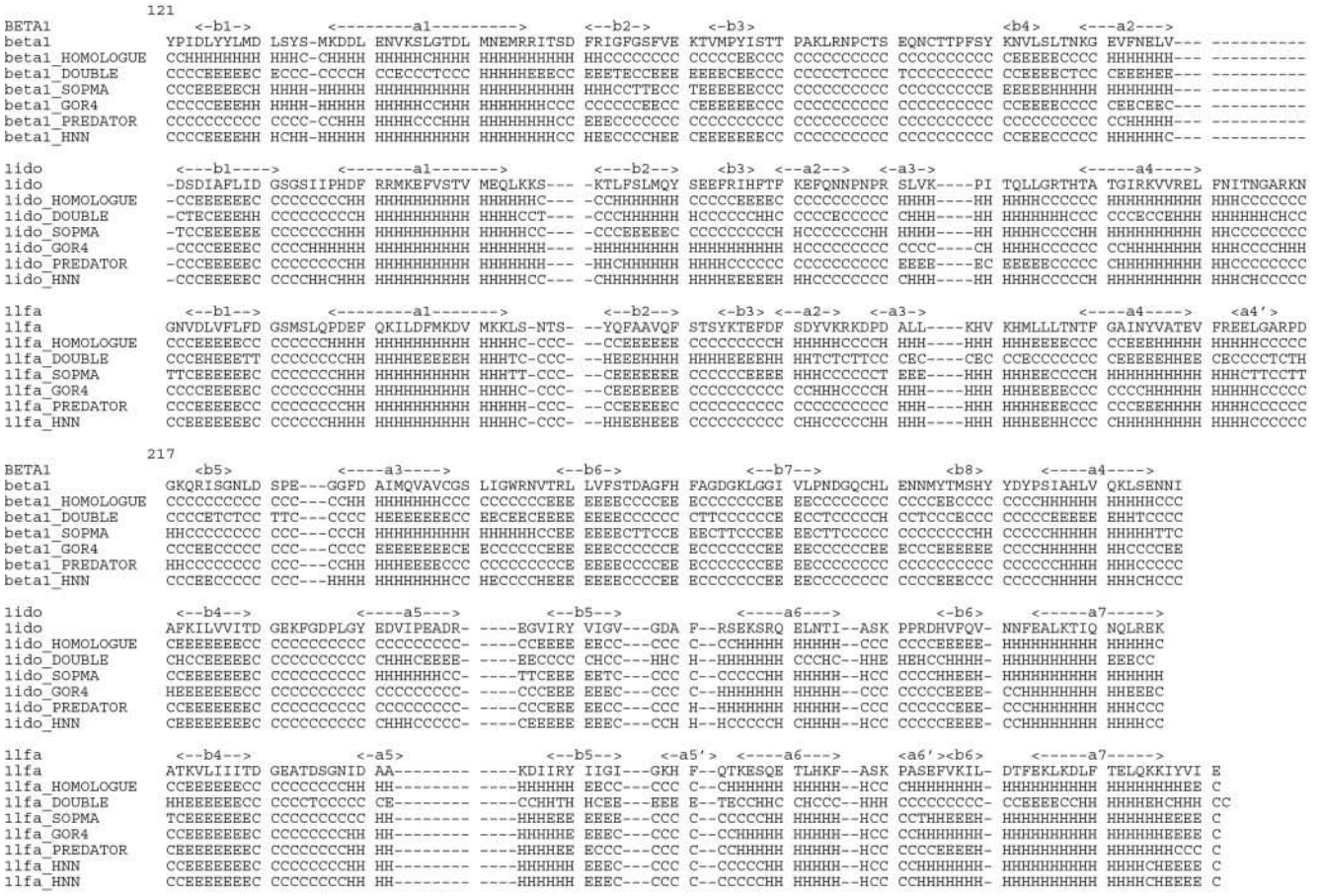

FIGURE 1 Results of secondary structure predictions on the β1, CD11A(1lfa), and CD11B(1ido) sequences. E, H, and C are character codes for residues predicted in β strand, α helix, and coil regions, respectively. The consensus-predicted α and β motifs are represented by a and b, respectively, followed by a number representing the count of that motif type. The motifs with a $'$ attached indicate that they occur in only one of the template proteins.

MATERIALS AND METHODS

General

The Homology and Discover modules (Biosym/MSI, 1995) of the Insight® II molecular modeling system were used in the construction and optimization of the β1 structure. The CVFF force field was used for all energy calculations, and nonbonded interactions were handled fully without using a cutoff. A distance-dependent dielectric constant of $4r$ was applied to all nonbonded interactions to simulate an aqueous environment. The sequence for β1 was obtained from National Center for Biotechnology Information (Wheeler et al., 2000) as the PIR accession number B27079 (Barker et al., 2000). Only a portion of β1 sequence from residues 121 to 311, which matches the length of the template protein, was used in this study.

Template selection

Because it is widely argued that β1 may share many biological features of the I domain that exists in a number of integrin α subunits (Puzon-McLaughlin and Takada, 1996; Tuckwell and Humphries, 1997), we used this as a primary guide for the selection of a template for β1. A detailed examination of the two-dimensional (2D) structural features of the I domains from a number of integrin α subunits whose crystallographic data are available indicated that the structure-conserved regions (SCRs) of the I domain are composed mainly of alternating α-helical and β-strand motifs. We then compared the consensus β1 secondary composition (2D motifs and their distribution) predicted by various protein secondary structure prediction programs (Deleage et al., 1997) with those present in CD11A/CD11B, two integrin I domains with the highest resolution crystal structures. Closely resembled secondary composition patterns were easily identified, as shown in Fig. 1. Therefore, we have chosen CD11A/CD11B as our template proteins. To validate our approach based on consensus 2D structure prediction, we also applied the same programs to CD11A/CD11B and compared the consensus-predicted 2D structures with those presented in their crystal structure. An extremely high fit can be identified for each of them. To confirm the choice of I domain as a suitable template for β1, we also conducted a series of protein fold recognition studies using various programs such as GENEFOLD (Sybyl, 1999). Results from these studies also strongly supported our template selection, because all programs gave a rather high preference for CD11A/CD11B.

Sequence alignment

Three steps were taken to achieve a highly optimized sequence alignment for β1. First, the consensus 2D structure pattern of β1 was compared to those of CD11A and CD11B, and adjustments were made to maximize the match of the corresponding 2D motifs, as shown in Fig. 2. Mapping of 2D elements ensures the inheritance of SCRs from the template protein. Second, for variable regions in the β1 sequence sharing low homology with

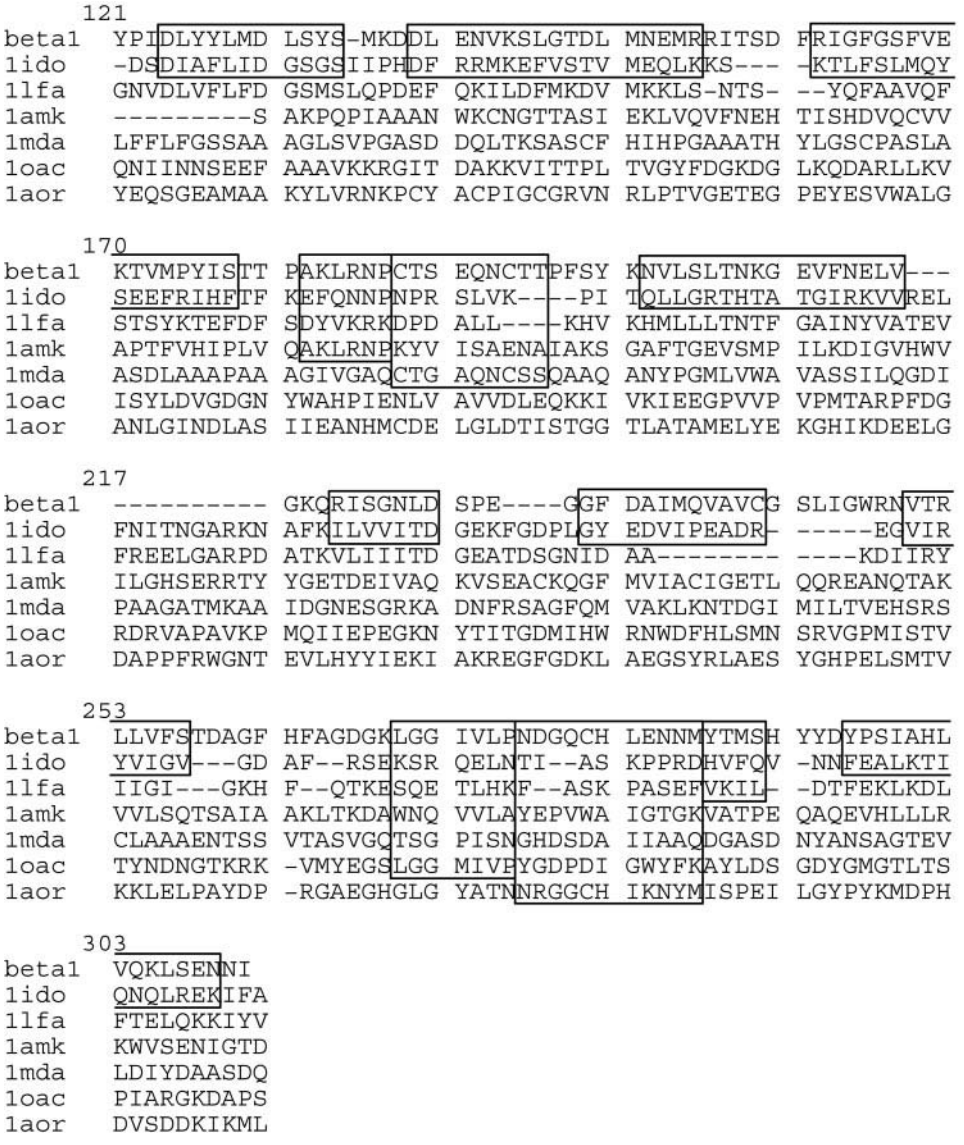

FIGURE 2 Alignment of CD11A, CD11B, and β1 sequences. Also listed are other protein sequences providing local templates. The sequence boxes cover the regions in which the backbone of a template is copied into the target protein. The notation for each PDB entry is given as follows: 1amk.pdb, Leishmania Mexicana triose phosphate isomer; 1mda.pdb Chain J, methylamine dehydrogenase (E.C.1.4.99.3) with amicyanin; 1oac.pdb, copper amine oxidase; 1aor.pdb, aldehyde ferredoxin oxidoreductase protein complexed with molybdopterin.

the template, a BLAST (Altschul et al., 1997) search was carried out to search for protein fragments with a higher sequence homology. If such fragments were found and a consensus configuration could be derived from them, then these fragments were used as a local template for that region. Third, final adjustments were made to align important residues, such as those in the binding site and those involved in disulfide linkages. At the same time, careful attention was paid to minimize inserted gaps and maximize the overall sequence homology.

Initial model generation

After the sequence alignment, two steps were taken to build an initial 3D model of β1 subunit. First, the backbone coordinates of the SCRs were copied from the template protein, and each side chain in the template was replaced by that of the corresponding target residue by aligning the C_α–C_β bonds. This is a standard procedure in most homology modeling packages. Second, the coordinates of the residues in variable regions were copied from peptide fragments generated by one of the following methods: a local template fragment if found, protein databank screening, and de novo generation. The last two methods are provided through the Homology module of Insight II package from MSI (Biosym/MSI, 1995). In the databank screening method, a search is done for real protein fragments fitting the distance specified by the two ending points. It usually suffers from a low hit rate, possibly a distorted junction geometry and ignorance of sequence homology when selecting fragments. De novo generation provides a universal method for generating loops and it can be used when other methods fail. However, it does give artificial peptide geometry that is difficult to correct in later stages of structure refinement. Actually, our approach of selecting a local template is an improvement to the database screening method that takes into account sequence homology when select-

ing fragments. In our approach, the fragment representing the consensus configuration found via a BLAST search was carefully spliced with the template structure to best bridge the gap. To accomplish this splicing, the loop fragment was manually aligned with the ends of the template and both parts were extended or contracted one residue at a time until an acceptable junction geometry was obtained. To build the variable loop regions, we first applied our local template approach, and coordinates were generated whenever high homologous fragments could be identified. Coordinates for the variable regions where the local templates could not be selected were assigned from either a protein databank screen or de novo generation if the prior two methods failed. Avoiding conflicts with other sections of the protein and fitting the overall template topology were primary criteria for determining a loop orientation for each variable region.

Structure refinement

A consequence of copying template coordinates in homology modeling is that the initial 3D structure inevitably contains serious conflicts among side chains of the constructed protein. This is due to changes in size and charge state. We designed a specific optimization protocol aimed at fully relaxing our initial model while preserving its inherited topological features. The refinement was accomplished in multiple steps involving restrained energy minimization and molecular dynamics. Once all atomic coordinates were assigned, the structure was initially refined by repairing junction distortions between loops and SCRs. It involved the minimization of residues spanning the two sides of a junction, and was accomplished with the tools available in the Insight II homology model (Biosym/MSI, 1995). Completing this step helps to correct the peptide bond length and torsion in the junctions and smooth out junction orientations. After this step, the site for placing Mg^{2+} was identified by overlapping the model and CD11B binding sites and copying the coordinates of Mg^{2+} from the template. Because all of metal coordinating residues in CD11A were matched to those in our sequence alignment, a reasonable assumption was made that the β1 model should share a similar metal coordination pattern with the template. To preserve the metal coordinating features, distance restraints were placed on all coordinating atom pairs during the following refinement steps. Because all Mg~O distances in the template metal coordination are between 2.00 and 2.25 Å, we used these two values as lower and upper boundaries of the distance restraints. In addition, the backbone atoms of all SCR residues were fixed during the initial stages of refinement process to prevent SCR from deforming significantly. First, a multiple-step steepest-decent (SD) minimization was conducted to correct most severe stereo conflicts and prepare the model for further relaxation. Then, a dynamics simulation was performed at 298 K to further release the stereo conflicts that survived from previous simple minimization protocol. At this point, the fixed atoms in SCRs hindered further relaxation, so the constraints on fixed backbone atoms in SCRs were replaced by tether restraints of 1000 kcal/(mol·Å^2) during the next multiple-step conjugate-gradient (CG) minimization process. The use of tether restraint was to preserve the overall geometric features of SCR regions. Last, a combined dynamics and minimization process with relaxed metal coordinating restraints was used to fine-tune the final structure model.

Ligand binding analysis

The native ligand VCAM and several synthetic compounds were used to study their possible binding modes to β1 and $\alpha 4\beta 1$ complex. First, for each ligand molecule whose conformation was undetermined, a thorough search using molecular dynamics of at least 100 ps was carried out, and the most energetically favorable conformation best fitting the pharmacophore model derived previously from active analog analysis (Holland, 1999) was chosen as its bioactive conformation. Second, each ligand molecule was manually introduced into the β1 active site to maximize the favorable interactions specified by the pharmacophore model. If such a favorable docking pattern could be identified, the system then underwent a further refinement protocol to determine the final binding mode. The protocol started with an initial minimization process of ligand with entire β1 fixed to adjust their relative orientation and correct any steric collision caused by manual docking. Next, a limited dynamics and minimization protocol was carried out on both ligand and β1 atoms in the active site (defined as residues within a spherical range of 12 Å from the metal ion) while keeping the rest of acceptor fixed. The last step of the refinement is necessary, because it allows both ligand and target to relax while searching for the relative orientation that optimizes the interactions between them.

RESULTS AND DISCUSSION

Sequence alignment

Because the overall sequence homology between β1 and the template proteins CD11A/CD11B is ~33.5%, we first focused on the alignment of the secondary structure motifs between them to ensure the preservation of SCRs. However, the sequence homology increases significantly (from 33.5% to 45.6%) as the local templates are applied to the sequence alignment as shown in Fig. 2. There were two helical regions present in both template proteins yet absent from predicted β1 secondary motifs. In one of these regions, the sequence fragment corresponding to two short helices, a2 and a3, in CD11A and CD11B was predicted as a random coil by all secondary structure prediction programs (See Fig. 1). Application of a BLAST search based on the local sequence in this region resulted in two highly homologous protein fragments, both of which were in coiled configurations. In the other region, the consensus secondary prediction indicated a β strand in a sequence segment where a helix (a6) was present in both CD11A and CD11B. Again, the BLAST search led us to a highly homologous protein fragment (5/10 identity, 9/10 similarity) in a β strand motif. Because both of these discrepancies in secondary motifs occurred in the surface regions of the template, it could be argued that helices in the template were less conserved than the β strands, which formed the topological core of I domain structure. Therefore, they might be more vulnerable to evolutionary mutations. This argument is further supported by the observation that the helical compositions in CD11A and CD11B were somewhat different, whereas those of β strands were rather consistent. For example, three short helices in CD11A (a4′, a5′, and a6′) were not present in CD11B. All of these arguments render ample support for using the protein fragments as local templates for these regions.

By our alignment, all Mg^{2+} coordinating residues in CD11B were aligned to the identical ones in β1 model, namely, D130, S132, S134, T206, and D226. Because most of these active site residues were determined by mutagenesis experiments as critical residues in ligand binding, (Puzon-McLaughlin and Takada, 1996) the residual identity inherited by our alignment would ensure the preservation of metal coordinating pattern of the template protein. Furthermore, our sequence alignment also easily accommodates

TABLE 1 Comparison of β1 with theoretical structures from the PDB

Value Calculated	Average β1 Structures (%)*	Average Theoretical (%)*
Core	75.8	75.2
Allowed	18.8	16.4
Generously allowed	3.6	2.3
Disallowed	1.8	6.2

The ϕ, ψ region percentages were calculated with PROCHECK V3.0. Only 176 out of the available 192 theoretical structures from PDB gave results with PROCHECK.

*Percentages were averaged from raw data values and were normalized to total 100%.

four cysteine residues that are absent from the template protein. Two of them are sequentially close to each other and sit in a single loop, whereas the other pair, though distant in sequence, is close enough in space to form a disulfide linkage.

Model evaluation

The structures of refined model were evaluated with the PROCHECK V3.0 (Laskowski et al., 1993). PROCHECK analyzes the backbone ϕ, ψ torsion angles of a protein and uses the percentage occupancies for "core," "allowed," "generously allowed," and "disallowed" regions as the quality measurement of a protein structure. For a fair assessment of our β1 model, we also conducted PROCHECK calculations on 176 out of 192 theoretical models (16 structures would not score properly as submitted to PROCHECK) available from the PDB databank. The values calculated for our β1 model and average values calculated for the 176 structures are shown in Table 1. The occupancy rates of both core and allowed regions for our β1 model are above the average values calculated from the theoretical structures, leading to a significant higher combined percentage in these favorable regions (94.6 versus 91.6%). In contrast, the percentage of residues in two less favorable regions is much lower for our model. By these comparisons, the quality of our β1 structure appears to be acceptable when compared to other theoretical models.

Figure 3 *A* shows a superposition of our β1 model with the template protein CD11B(1ido). It is clear that the overall topology of the template protein has been inherited while some significant local configuration changes are presented in variable regions. The Mg^{2+} coordinating pattern remains, although the sizes of the loops around the active site have been altered. A detailed examination of residual type reveals that the majority of charged residues are located on the surface of our model whereas the core region is made of mainly hydrophobic residues. This distribution, as portrayed in Fig. 3 *B*, has been confirmed by a solvent-exposed surface calculation in which the solvent exposed surface area was compared to the total surface area for each residue type. This led to an average ratio of ~0.7 for all polar residues and ~0.3 for all hydrophobic residues. There are four cysteine residues in β1 sequence that are not present in the template, and they may contribute significantly to the overall topology of our β1 model. According to our sequence alignment, one of them, C241, is located at the end of helix a3 and in close proximity to C281 in a neighboring loop between b7 and b8. Both of the remaining cysteines, C187 and C193, are in a single loop between b3 and b4 on the other side of the molecule. Therefore, our β1 model may have none, one, or two disulfide linkages. Generally speaking, the cysteines in a protein have a strong tendency to form disulfide linkages under favorable conditions, so we have chosen to adopt the disulfide bond between C241 and C281. Forming a disulfide linkage between C187 and C193 in the same loop, however, would result in a kink loop configuration due to their sequential closeness. Furthermore, as shown in Fig. 2, we identified a highly homologous protein fragment from methylamine dehydrogenase (PDB entry 1mda) in which the two cysteines in the fragment corresponding to the loop do not form a disulfide bond. Therefore, we chose not to make the disulfide linkage between C187 and C193, and instead model the loop from the local template.

Due to its biological importance, the β1 subunit has attracted an increasing number of experimental investigations aimed at elucidating the mechanism of ligand binding (Tuckwell et al., 1994; Chen et al., 1998; Puzon-McLaughlin and Takada, 1996). A number of critical residues in ligand binding were identified by site-directed mutagenesis technique. In addition to D130, which had previously been shown important for ligand binding, Puzon-McLaughlin also pin-pointed other residues, namely S132, N224, D226, E229, D233, D267, and D295, that might also be critical to ligand binding (Puzon-McLaughlin and Takada, 1996). It is interesting that, together with D130, S134 and T206, S132 and D226 were also aligned in our model to the corresponding MIDAS domain residues of the template proteins. As pointed out earlier, the conserved DXSXS sequence is believed to be involved in the metal ion coordination and integrin binding. Due to a high homology between the model and template sequences in this region, it is likely that these five residues of β1 are placed in topologically similar positions to that of metal coordinating residues in the template. Furthermore, due to residues N224 and E229 being located in the same loop region between b5 and a3 in our model and spatially close to metal coordinating residues (Fig. 3 *C*), they should also contribute directly to ligand binding. The other three residues, D233, D267, and D295, are located in the separated regions on the molecule surface (Fig. 3 *C*) and their involvement in ligand binding is not well explained by our stand-alone β1 model. Because all of them are located in the loop regions distant from the active site, the mutations performed on them may induce adverse effects on α/β association on immunoprecipitation under

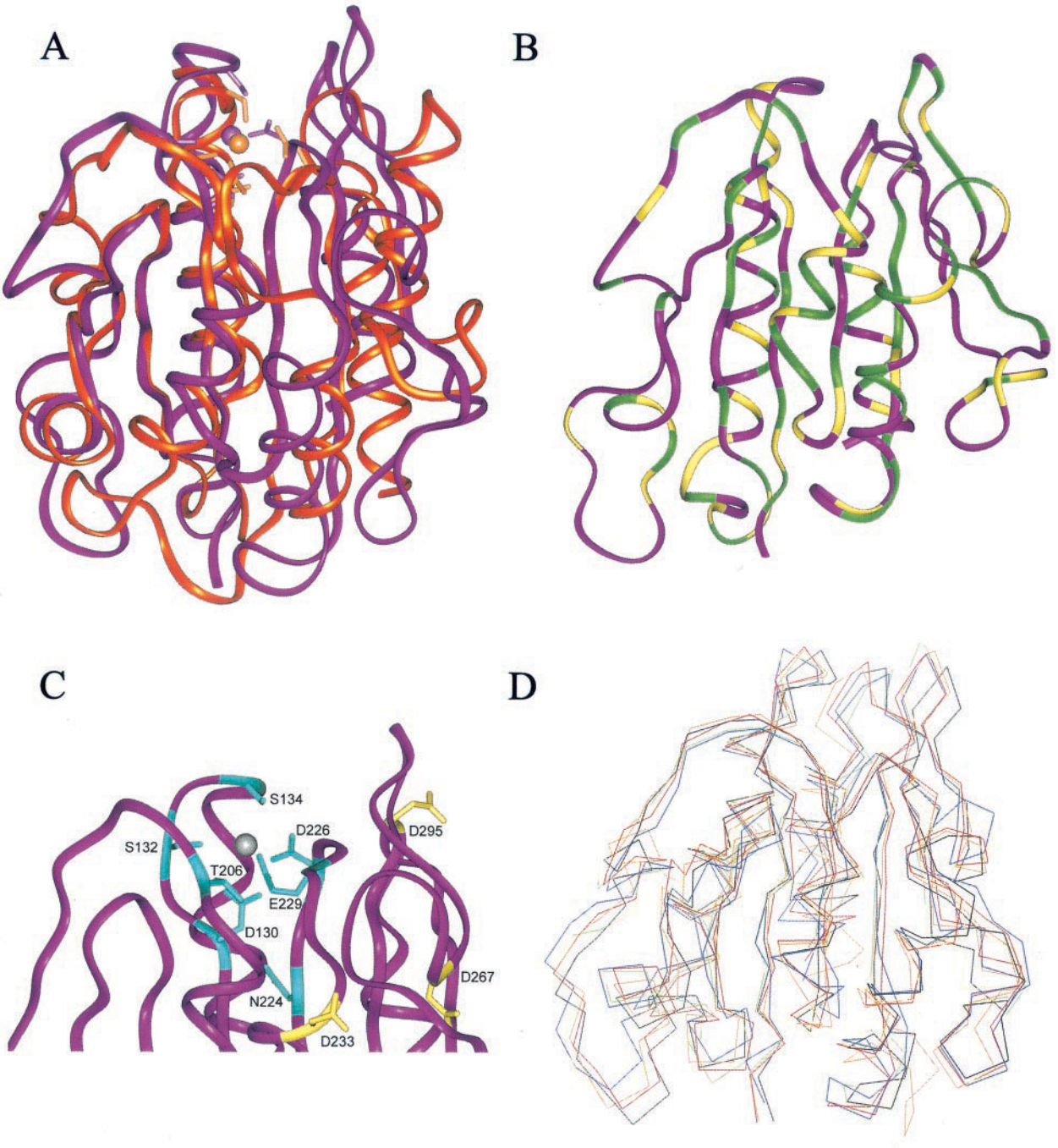

FIGURE 3 Color ribbon/trace representation of β1 model. (*A*) Comparison of refined β1 model in magenta with its template protein CD11B(1ido) in orange. Those residues involved in metal-ion coordination are highlighted. (*B*) Color-coded residue distribution of β1 model. Green, mostly hydrophobic residues (A, V, F, L, I); yellow, charged hydrophilic residues (D, E, K, R, H); purple, remaining. (*C*) Distribution of critical residues identified by the site directed mutagenesis data. Residues in blue are those believed to associate with metal-ion coordination, whereas those in yellow are speculated to contribute to $\alpha\beta$ association. (*D*) Superposition of β1 conformations extracted from a dynamics simulation. An ensemble of C_α traces in different colors represent snap shots of β1 conformations from a dynamic trajectory.

the experimental condition, as speculated by Puzon-McLaughlin and Takada (1996). Furthermore, their functional roles in ligand binding may become clear upon the association of β1 with α4 subunit (also see the VCAM binding analysis in the next section).

To observe the conformational flexibility of our β1 model, we conducted a limited trajectory analysis on an ensemble of minimized structures abstracted from a dynamic simulation. The molecular dynamics was carried out under room temperature without any restraints for 200 ps.

Figure 3 *D* shows the superposition of an ensemble of minimized β1 conformations abstracted from the dynamic trajectory. The standard deviation data suggest that the configurations of SCRs are rather preserved, whereas the loop regions on both top and bottom of the model demonstrate a certain degree of conformational mobility. This is in line with our expectation, because the SCRs mainly form the structure core of our β1 model and, therefore, are spatially more constrained.

Ligand binding analysis

The refined β1 model was used to study its ligand-binding mechanism. The first molecule used is an active cyclic peptide TBC772 derived from a section within the CS-1 splice variant of fibronectin (McIntyre et al., 1997). Cyclization of this short peptide greatly reduced the backbone conformational flexibility. Figure 4 *A* shows the detailed binding interactions between TBC772 and active site residues in our β1 model. The carboxylic acid of Aspartic acid in TBC772 lies above the Mg^{2+} ion in its coordination sphere. This coordination is supported by the crystal structure of CD11B, which shows the sixth coordination site occupied by residue E314 from a neighboring CD11B molecule (Lee et al., 1995a,b). The Val and Leu in TBC772 share favorable hydrophobic interactions with L205 and P228 of the β1 model, respectively. Furthermore, the NH of the indole ring in tryptophan can form a hydrogen bond with E229 in our β1 model. This interaction might partially account for the role of E229 in ligand binding. Given the intrinsic flexibility of a homology model, one might argue that L205 and P228, located in two different loops around an active site, are, by chance, placed near the active site during our model building. However, this possibility can be eliminated if we examine their sequential relationship with the key residues for metal ion coordination. For example, L205 is next to T206 and P228 is sandwiched between D226 and E229. Therefore, the locations of these two residues are rather restrained in the active site due to the conserved MIDAS motif geometry. Furthermore, both L205 and P228 are evolutionarily conserved throughout all β subunits, implicating possible biological functions. From the above discussion, it is clear that each of four peptide residues in TBC772 has found its favorable partner in the active site, and these interactions should have contributed to the high binding affinity of TBC772 for β1.

With the success of docking TBC772 to our β1 model, we extended the ligand binding study to include some active nonpeptide compounds. One of these compounds, TBC3486, is a VLA-4 antagonist with an IC50 of 0.4 nM in the CS-1 assay. Figure 4 *B* presents a possible binding mode of TBC3486 to our β1 model. This binding mode was achieved by superimposing its carboxylic acid with that of docked TBC772 while maximizing the contact between its two hydrophobic groups with L205 and P228 of β1. According to this binding mode, S227 and K208 in β1 are hydrogen bonded, respectively, to the urea NH groups and the nonurea carbonyl oxygen of TBC3486, two critical pharmacophore features determined from our active analog analysis (Holland, 1999). Also, Y133 is hydrogen bonded to one of the oxygens in the methylenedioxy fragment of TBC3486. Residues Y133 and S227 are another pair of interesting active site residues, because, with a hydroxyl group, they can act as either a hydrogen-bond donor or acceptor, facilitating their hydrogen-bonding capacity with ligands. For instance, in addition to its hydrogen bond to the urea NH groups in TBC3486, S227 can also form a hydrogen bond with the carboxylic acid group of E229 and stabilize the configuration of the loop. In contrast, residue Y133 may also contribute to hydrophobic interaction between ligand and receptor by ring stacking. Arguments for this type of interaction are supported by the results of a binding assay in which there was no apparent loss in activity after replacing methylenedioxyl in TBC3486 with a substituted aromatic ring (unpublished data). Like L205 and P228, the positions of Y133, K208, and S227 in the active site are rather limited because they are also neighbored by the key metal ion-coordinating residues. However, these three residues are not evolutionarily conserved throughout β subunits, and their participation in ligand binding might lead to binding selectivity among different β subunits.

The binding analysis of small synthetic compounds presented above provides us rich information about the mechanism by which β1 interacts with its ligands. This in turn facilitates our ability to understand how α4β1 complex interacts with its native ligand VCAM. It has been shown that residues QIDSPY in domain I of VCAM are critical in its binding to α4β1 and, by symmetry, a highly charged loop in domain II is also speculated to be important in binding (Newham et al., 1997; Green et al., 1999). It is interesting that IDSP loop shares sequence similarity with TBC772, which was derived from CS-1. This suggests that it might bind to β1 in a similar fashion to TBC772. Docking of this loop into β1 active site by aligning the carboxylic acid of Asp to the same group in docked TBC772 (see Fig. 4 *C*) leads us to the orientation shown in Fig. 4 *D*. It is interesting that this docking orientation retains most of favorable interactions between the ligand and receptor while avoiding steric clashes between the domains of two molecules. Again, L205 and P228 share close hydrophobic interactions with proline and isoleucine in the VCAM loop, respectively, and the OH group of serine in the loop is within the hydrogen bonding range of both N207 and E229 of β1 model. An interesting observation of this docking orientation is that it places the other domain of VCAM in such a position that its highly charged loop is in the neighborhood of β1 residues D233, D267, and D295. One side of this loop, LEDADRKSL, is arrayed by positively charged residues with hydrogen-bond donor potential, whereas another side is made of negatively charged hydrogen-bond

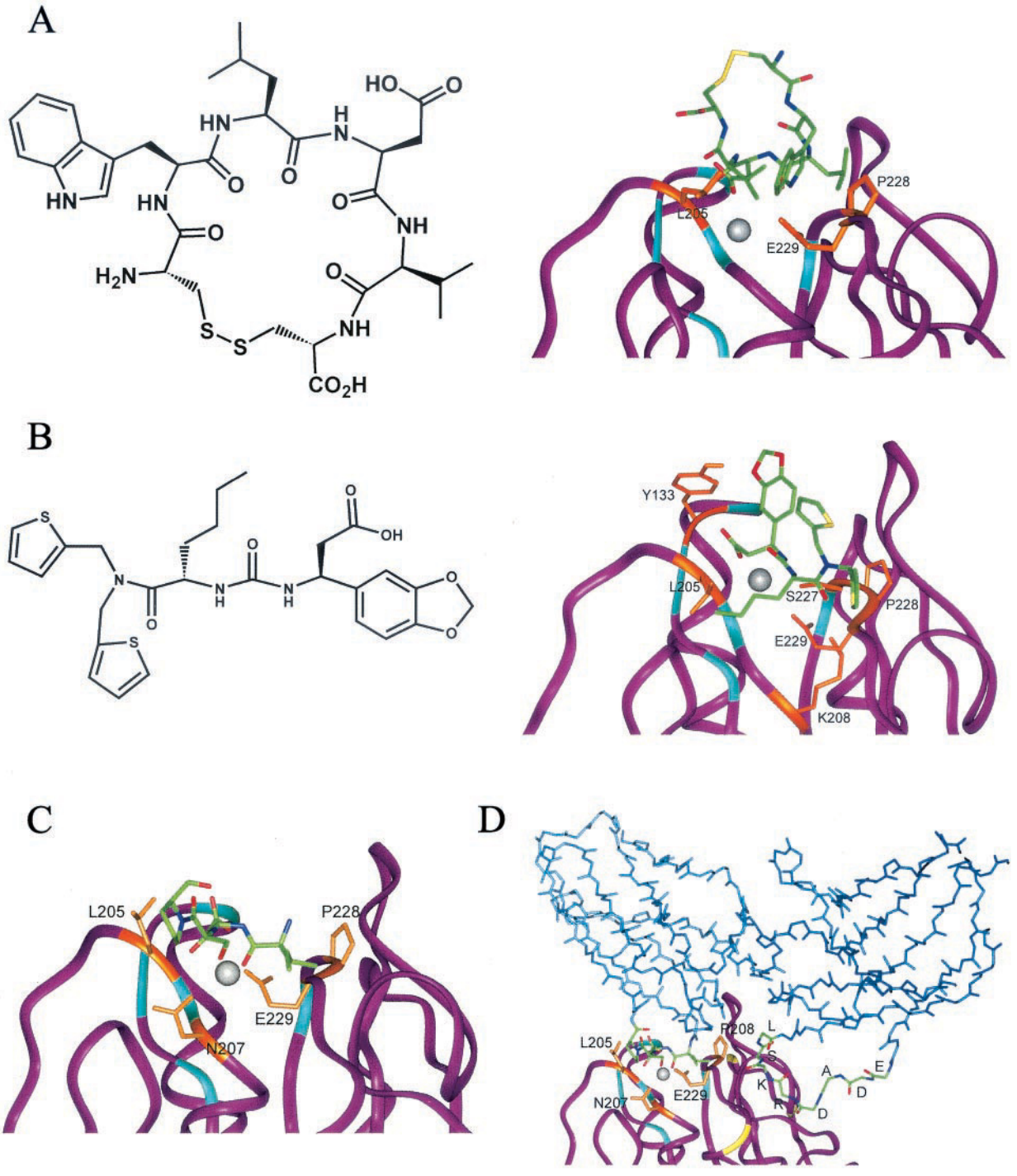

FIGURE 4 Ligand-binding analysis based on the refined β1 model. Ligands are colored by their atom types: O = red; N = blue; C = green; S = yellow. β1 residues proposed as critical for ligand binding are indicated by orange, whereas those in blue denote metal-coordinating residues. (*A*) Docking of TBC772 to β1 model. A 2D sketch of TBC772 structure is also included. (*B*) Docking of TBC3486 to β1 model. A 2D sketch of TBC3486 is also attached. (*C*) Docking of the IDSP loop in VCAM domain I to β1. (*D*) Docking of VCAM domain 1 to β1 model leads to a specific orientation of the LEDADRKSL loop in VCAM domain II to β1 model. VCAM is colored light blue and three critical binding residues in β1, D233, D267, and D295, are colored yellow.

acceptor residues. Docking of VCAM domain I loop orients the positively charged portion of the loop toward key β1 residues D233 and D267 and places the negatively charged section exposed to the space that would possibly be occupied by the α subunit. This would suggest that the binding site of α4 subunit should be either metal-ion coordinated or

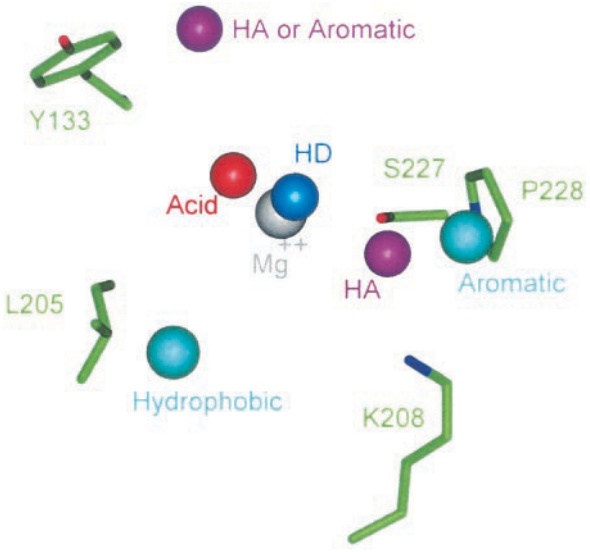

FIGURE 5 Comparison of the pharmacophore features derived from active analog analysis to β1 active site residue distribution. Active site residues proposed to be critical for binding are colored by their atom types, and the metal ion Mg^{2+} is represented by a gray sphere. Each pharmacophore feature is also represented by a sphere of different colors: Orange, acid; purple, hydrogen bond acceptor (HA); blue, hydrogen bond donor (HD); and green, hydrophobic group.

in a highly positively charged state. This is consistent with a model of the α subunit in which a Mg ion is predicted to bind to its upper face (Springer, 1997). With the proposed involvement of residues Y133, L205, K208, S227, and P228 to the binding of TBC772/TBC3486 type ligands to β1 subunit, it is useful to compare their distribution in the active site to a previously derived pharmacophore model (Holland, 1999). In Fig. 5, the geometric relationship between the key binding residues in our β1 model is compared to the pharmacophore model. It is important to understand that the pharmacophore, which specifies the geometric relationship between the key functional groups in active compounds, should be complementary to the characteristics of critical residues in the active site. Obviously, L205 and P228 can fill two hydrophobic regions specified by the pharmacophore model rather well. Residues Y133, K208, and S227 can each map to one of the hydrogen bonding features in the pharmacophore model, namely Y133 to HA or aromatic, K208 to HD, and S227 to HA. In matching both the features and the geometry specified by the pharmacophore model, our β1 model provides us with a receptor structure-based explanation on why the functional elements of the pharmacophore model are important to ligand binding.

β1/β7 Binding selectivity

Selectivity between α4β1 and α4β7 is important because the diseases attenuated by inhibition of these homologous integrins are different in many cases (Green et al., 1999). Our activity assay results for TBC3486 show a 1000-fold selectivity in favor of α4β1 (unpublished data). Because the sequence homology is extremely high for these two β subunits (~61% identity and ~78% similarity) they should share closely resembled structural and topological features (see Fig. 6 *A*). It is reasonable to assume that the binding selectivity between these two subunits arises from the mutations of some critical residues, especially those in the active site. However, the sequence alignment indicates that most key residues for metal-ion coordinating and ligand binding, including those suggested by mutagenesis experiments and those proposed by our model, are totally conserved between β1 and β7. Therefore, those residues seem unlikely to account for binding selectivity. In search of the residues that might be responsible for the selectivity, we investigated all residues within 12 Å of the Mg ion in our β1 model, which effectively covers the entire β1 active site. All residues in this region except five are conserved between β1 and β7. Out of these five, only Asn/Gly-207 and Lys/Asp-208 may cause significant physiochemical disturbance to their local area. A detailed examination of the structural environment of these two residues in our β1 model has revealed the structural roles these mutations might play in the ligand binding selectivity. According to our β1 model, N207 and K208 are located in the loop between b4 and a2 next to the metal ion coordinating residue T206. As shown in Fig. 6 *B*, the positively charged K208 has found itself a comfortable electrostatic environment in which it is surrounded by the E229 side chain and the backbone carbonyls of S227 and P228. Residue N207 also forms a hydrogen bond with the carboxylic acid group of E229. As discussed earlier, E229 was identified by mutagenesis experiments as critical for β1 ligand binding and it is sequentially next to residues D226–P228, and together they form a loop between b5 and a3. The favorable electrostatic interaction between N207K208 and E229 should play an important role in achieving and maintaining the relative orientation between these two loops in our β1 model. The Asn/Gly-207 and Lys/Asp-208 changes not only cause the loss of the hydrogen bond between N207 and E229, but also introduce a large electrostatic repelling force between D208 and E229. These two modifications cause a reorientation between the two loops. Since both loops contain a number of critical residues in metal coordination and ligand binding their reorientation will inevitably alter the geometric distribution of these residues and this is presumably the cause of ligand selectivity between β1 and β7. To lend support to this argument, we have conducted a limited loop conformational search on a modified β1 model in which N207 and K208 were replaced by G207 and D208, respectively. The search was carried out through a restrained dynamics simulation in which all residues other than those in the active site were fixed. The conformations sampled from the dynamics trajectory were then refined by a minimization protocol. In Fig. 6 *C*, the active site loops in our original β1 model are

A

```
          1
beta1  YPIDLYYLMD LSYSMKDDLE NVKSLGTDLM NEMRRITSDF RIGFGSFVEK
beta7  YPVDLYYLMD LSYSMKDDLE RVRQLGHALL VRLQEVTHSV RIGFGSFVDK

         51
beta1  TVMPYISTTP AKLRNPCTSE -QNCTTPFSY KNVLSLTNKG EVFNELVGKQ
beta7  TVLPFVSTVP SKLRHPCPTR LERCQSPFSF HHVLSLTGDA QAFEREVGRQ

        101
beta1  RISGNLDSPE GGFDAIMQVA VCGSLIGWRN VTRLLVFSTD AGFHFAGDGK
beta7  SVSGNLDSPE GGFDAILQAA LCQEQIGWRN VSRLLVFTSD DTFHTAGDGK

        151
beta1  LGGIVLPNDG QCHLENN-MY TMSHYYDYPS IAHLVQKLSE NNI
beta7  LGGIFMPSDG HCHLDSNGLY SRSTEFDYPS VGQVAQALSA ANI
```

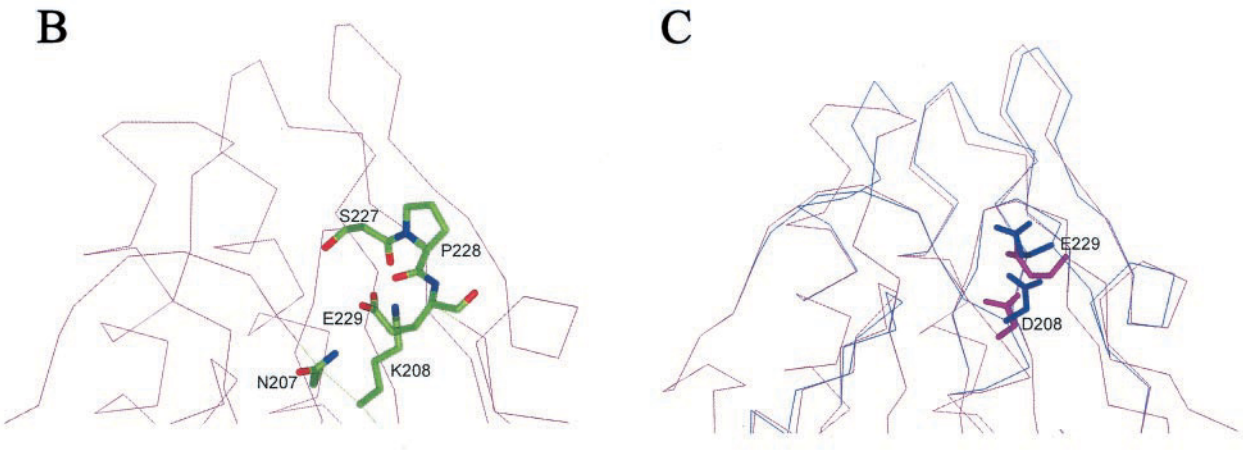

FIGURE 6 Proposed mechanism of ligand binding selectivity between β1 and β7. (*A*) Sequence alignment between β1 and β7. (*B*) The local environment of residue K208 in β1 model. Critical residues are colored by their atom types. (*C*) The simulated local conformational changes caused by Asn/Gly-207 and Lys/Asp-208 from β1 to β7. Original β1 model is colored purple, and modified β1 model is colored blue.

compared to those in derived β7 model with special highlight to the positions of K208D and E229. It is clear that D208 and E229 have moved significantly from their original positions in β1 to compensate the unfavorable electrostatic interaction between them. The reposition of these two side chains causes loops in the active site to reorient, which in turn alters the positions of many critical binding residues. Though it is beyond the scope of this paper, a quantitative description of change in binding affinity, such as those provided by a free energy calculation, should provided more convincing support. Nevertheless, the analysis presented here provides a qualitative explanation for binding selectivity between β1 and β7 subunits.

CONCLUSIONS

Based on the template of I domain from integrin α subunit CD11A/CD11B, we have built a 3D structural model for β1 subunit and used the model to study the possible binding mechanism of β1 with VCAM and other highly active ligands. The sequence alignment based on the mapping of secondary structure motifs ensured the preservation of structurally conserved regions from the template. The refined β1 structure is comparable in quality to other theoretical models and the model has a reasonable residual distribution in which most hydrophilic residues are located on the molecular surface. Furthermore, the model can satisfactorily explain most mutagenesis data available. Based on the ligand docking analysis, the model active site fits our previously derived pharmacophore specification (Holland, 1999) very well, and a number of residues that might also contribute significantly to ligand binding have been identified. The model has also been used to propose the possible mechanism behind β1/β7 ligand binding selectivity. In summary, the model provides valuable structure information about the ligand binding to the β1 subunit and α4β1complex, which in turn benefits the search for potential therapeutic agents.

The authors feel grateful to Peter Vanderslice for his critical reading of the manuscript and helpful discussion on $\beta1/\beta7$ ligand binding selectivities. Special thanks go to George Krudy and Jeffery Luo for their technical assistance in preparing data used in this manuscript.

REFERENCES

Adams, S., and R. Lobb. 1999. Inhibitors of integrin alpha 4 beta1 (VLA-4). *In* Annual Reports in Medicinal Chemistry. W. Hagmann, editor. Chap. 18. Academic Press, New York. 179–188.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402.

Barker, W. C., J. S. Garavelli, H. Huang, P. B. McGarvey, B. C. Orcutt, G. Y. Srinivasarao, C. Xiao, L. S. Yeh, R. S. Ledley, J. F. Janda, F. Pfeiffer, H. W. Mewes, A. Tsugita, and C. Wu. 2000. The protein information resource (PIR). *Nucleic Acids Res.* 28:41–44.

Biosym/MSI. 1995. Homology User Guide. Biosym/MSI, San Diego, CA.

Chen, L., R. Lobb, J. Cuervo, K. Lin, S. Adams, and R. Pepinsky. 1998. Identification of ligand binding sites on integrin alpha4beta1 through chemical cross-linking. *Biochemistry.* 37:8743–8753.

Chothia, C., and E. Y. Jones. 1997. The molecular structure of cell adhesion molecules. *Annu. Rev. Biochem.* 66:823–862.

Deleage, G., C. Blanchet, and C. Geourjon. 1997. Protein structure prediction: implications for the biologist. *Biochimie.* 79:681–686.

Elices, M., L. Osborn, Y. Takada, C. Crouse, S. Luhowskyj, M. Hemler, and R. Lobb. 1990. VCAM-1 on activated endothelium interacts with the leukocyte integrin VLA-4 at a site distinct from the VLA-4/fibronectin binding site. *Cell.* 60:577–584.

Emsley, J., S. King, J. Bergelson, and R. Liddington. 1997. Crystal structure of the I domain from integrin alpha2beta1. *J. Biol. Chem.* 272: 28512–28517.

Green, N., J. Rosebrook, N. Cochran, K. Tan, J.-H. Wang, T. Springer, and M. Briskin. 1999. Mutational analysis of MAdCAM-1/alpha4 beta7 interactions reveals significant binding determinants in both the first and second immunoglobulin domains. *Cell Adhes. Commun.* 7:167–181.

Haas, T., and E. Plow. 1994. Integrin–ligand interactions: a year in review. *Curr. Opin. Cell Biol.* 6:656–662.

Holland, G. 1999. Novel and highly potent and selective VLA-4 antagonists. *In* Abstracts of Papers, 219th ACS National Meeting, San Francisco, CA. American Chemical Society, Washington, DC.

Horwitz, A. 1997. Integrins and health. *Sci. Am.* 275:68–75.

Irie, A., T. Kamata, W. Puzon-McLaughlin, and Y. Takada. 1995. Critical amino acid residues for ligand binding are clustered in a predicted beta-turn of the third N-terminal repeat in the integrin alpha 4 and alpha 5 subunits. *EMBO J.* 14:5550–5556.

Jackson, D. Y., C. Quan, D. R. Artis, T. Rawson, B. Blackburn, M. Struble, G. Fitzgerald, K. Chan, S. Mullins, J. P. Burnier, W. J. Fairbrother, K. Clark, M. Berisini, H. Chui, M. Renz, S. Jones, and S. Fong. 1997. Potent alpha 4 beta 1 peptide antagonists as potential anti-inflammatory agents. *J. Med. Chem.* 40:3359–3368.

Kamata, T., W. Puzon, and Y. Takada. 1995. Identification of putative ligand-binding sites of the integrin alpha 4 beta 1 (VLA-4, cd49d/CD29). *Biochem. J.* 305:945–951. Published erratum: *Biochem. J.* 1996; 317:959.

Kc, L., H. S. Ateeq, S. H. Hsiung, L. T. Chong, C. N. Zimmerman, A. Castro, W. C. Lee, C. E. Hammond, S. Kalkunte, L. L. Chen, R. B. Pepinsky, D. R. Leone, A. G. Sprague, W. M. Abraham, A. Gill, R. R. Lobb, and S. P. Adams. 1999. Selective, tight-binding inhibitors of integrin alpha4beta1 that inhibit allergic airway responses. *J. Med. Chem.* 42:920–934.

Laskowski, R., M. MacArthur, D. Moss, and J. Thorton. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 26:283–291.

Lee, J., L. Bankston, M. Arnaout, and R. Liddington. 1995a. Two conformations of the integrin A-domain (I-domain): a pathway for activation? *Structure.* 3:1333–1340.

Lee, J., P. Rieu, M. Arnaout, and R. Liddington. 1995b. Crystal structure of the A domain from the alpha subunit of integrin CR3 (CD11b/CD18). *Cell.* 80:631–638.

Loftus, J., J. Smith, and M. Ginsberg. 1994. Integrin-mediated cell adhesion: the extracellular face. *J. Biol. Chem.* 269:25235–25238.

McIntyre, B., D. Woodside, D. Caruso, D. Wooten, S. Simon, S. Neelamegham, J. Revelle, and P. Vanderslice. 1997. Regulation of human T lymphocyte coactivation with an alpha4 integrin antagonist peptide. *J. Immunol.* 158:4180–4186.

Newham, P., S. Craig, G. Seddon, N. Schofield, A. Rees, R. Edwards, E. Jones, and M. Humphries. 1997. Alpha4 integrin binding interfaces on VCAM-1 and MAdCAM-1: integrin binding footprints identify accessory binding sites that play a role in integrin specificity. *J. Biol. Chem.* 272:19429–19440.

Puzon-McLaughlin, W., and Y. Takada. 1996. Critical residues for ligand binding in an I domain-like structure of the integrin beta1 subunit. *J. Biol. Chem.* 271:20438–20443.

Schiffer, S., M. Hemler, R. Lobb, R. Tizard, and L. Osborn. 1995. Molecular mapping of functional antibody binding sites of alpha 4 integrin. *J. Biol. Chem.* 270:14270–14273.

Springer, T. 1997. Folding of the N-terminal, ligand-binding region of integrin alpha-subunits into a beta-propeller domain. *Proc. Natl. Acad. Sci. U.S.A.* 94:65–72.

Sybyl. 1999. SYBYL, Version 6.6. Tripos, Inc., St. Louis, MO.

Tuckwell, D., and M. Humphries. 1997. A structure prediction for the ligand-binding region of the integrin beta subunit: evidence for the presence of a von Willebrand factor A domain. *FEBS Lett.* 400:297–303.

Tuckwell, D., M. Humphries, and A. Brass. 1994. A secondary structure model of the integrin alpha subunit N-terminal domain based on analysis of multiple alignments. *Cell Adhes. Commun.* 2:385–402.

Vanderslice, P., K. Ren, J. Revelle, D. Kim, D. Scott, R. Bjercke, E. Yeh, P. Beck, and T. Kogan. 1997. A cyclic hexapeptide is a potent antagonist of alpha 4 integrins. *J. Immunol.* 158:1710–1718.

Wang, J., R. Pepinsky, T. Stehle, J. Liu, M. Karpusas, B. Browning, and L. Osborn. 1995. The crystal structure of an N-terminal two-domain fragment of vascular cell adhesion molecule 1 (VCAM-1): a cyclic peptide based on the domain 1 C-D loop can inhibit VCAM-1-alpha 4 integrin interaction. *Proc. Natl. Acad. Sci. U.S.A.* 92:5714–5718.

Wayner, E., A. Garcia-Pardo, M. Humphries, J. McDonald, and W. Carter. 1989. Identification and characterization of the T lymphocyte adhesion receptor for an alternative cell attachment domain (CS-1) in plasma fibronectin. *J. Cell Biol.* 109:1321–1330.

Wheeler, D. L., C. Chappey, A. E. Lash, D. D. Leipe, T. L. Madden, G. D. Schuler, T. A. Tatusova, and B. A. Rapp. 2000. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 28:10–14.