Structure *16*

Supplemental Data

Integrin Conformational Regulation: Uncoupling Extension/Tail Separation from Changes in the Head Region by a Multiresolution Approach

Mattia Rocco, Camillo Rosano, John W. Weisel, David A. Horita, and Roy R. Hantgan

Supplemental Results

Completing the $\alpha_v\beta_3$ Ectodomain Model

We started from the PDB structure 1U8C (Xiong et al., 2004) in which residues 839-867 and 957 in α_v and 51-53, 435-522, and 691-692 in β_3 are not resolved, although present in the expressed construct (Adair et al., 2005). The α839-867 stretch, located in the calf2 module EE' loop (which contains the proteolytic cleavage site after Arg860 generating the heavy and light chains in mature $\alpha_v\beta_3$), was approximately modeled. The $\beta_3$435-522 stretch, corresponding to the I-EGF1/I-EGF2 modules, was initially homology-modeled using the β_2 I-EGF3 NMR structure (Beglova et al., 2002) and positioned as reported for $\alpha_{IIb}\beta_3$ (Hantgan et al., 2003). The remaining β_3 51-53 residues belonging to the PSI module were then modeled to complete our basic $\alpha_v\beta_3$ ectodomain model (Av-1). Since the original $\alpha_v\beta_3$ construct was expressed in baculovirus (Mehta et al., 1998), it is probably glycosylated with high-mannose-type carbohydrates. A prototypical $[GlcNac]_2[Man]_{11}$ chain was built from structures kindly provided by R.J. Woods (Complex Carbohydrate Research Center, University of Georgia, Athens, GA) and attached to 8 Asn residues in α_v (44, 260, 266, 458, 585, 821, 943, 950) and 5 in β_3 (99, 320, 452, 559, 654) in the Av-1 model, generating model Av-2. Subsequently, we made a homology model of the PSI/I-EGF1/2 modules based on the recently published β_2 PSI/hybrid/I-EGF1-2 crystal structure 2P26 (Shi et al., 2007) and replaced it in the Av-2 model. The I-EGF2/I-EGF3 interface was adjusted to conform to that present in the β_2 PSI/hybrid/I-EGF1-3 crystal structure 2P28 published together with 2P26 (Shi et al., 2007), resulting in a first alternative $\alpha_v\beta_3$ ectodomain model (Av-3).

Generating the Partially Opened $\alpha_v\beta_3$ Ectodomain Model

Model Av-6/N20 was built starting from model Av-4/S10, using a thigh/calf1 structure with a 20° unbending and keeping fixed the N-terminal domains of both chains down to the PSI/hybrid domain in the β_3 chain. Then the C-terminal parts of both chains (up to calf1 in α_v and I-EGF3 in β_3) were rigidly attached, superimposing the two calf1 modules. The I-EGF1-2 modules were inserted, again by superimposing the new PSI/hybrid/I-EGF1-2 homology model based on the 2P26 crystal structure (Shi et al., 2007), using the hybrid domain as a template. Their orientation was slightly changed to bring them more in line with the C-terminal end of the chain, using mainly Glu55 and Asp434 as pivots, and respecting all the original disulfide bridges. To maintain the long-range Cys55-Cys235 bridge, the orientation of the PSI module was also slightly changed. The I-EGF3 module was then reconnected at Glu522, which therefore acted as the main pivot for the conformational change, and the interface between the I-EGF3/4 modules was adjusted at Asn559, again keeping all the disulfide bridges intact.

Building a Complete $\alpha_{IIb}\beta_3$ Model

We previously reported (Hantgan et al., 2003, 2004) generating models for $\alpha_{IIb}\beta_3$ based on the $\alpha_v\beta_3$ crystal structure (Xiong et al., 2001) and showed that the fully bent form is at odds with solution data (Hantgan et al., 2004). Here, we used the Av-6/N20 model as the template for the extracellular region and added the resting TM helices model developed by Gottschalk (Gottschalk and Kessler, 2004; Gottschalk, 2005; kindly provided by K.E. Gottschalk, Weizmann Institute of Science, Rehovot, Israel) and our previously developed models of the cytoplasmic regions (Hantgan et al., 2003) based on NMR data (Vinogradova et al., 2000; Ulmer et al., 2001). N-linked complex carbohydrate models (see Rai et al., 2005) were attached to previously identified (Rocco et al., 1993) Asn residues. O-linked carbohydrate ($[Gal]_1[GalNac]_1[NeuNAc]_1$) models were similarly developed and attached, based on the prediction of NetOGlyc 2.0 (Julenius et al., 2005), to Thr or Ser α_{IIb} residues at seven newly identified locations (42, 259, 413, 576, 619, 845, 847). Fifty OG molecules were positioned around the TM helices to mimic the solubilizing micelle (Hantgan et al., 2003).

Reanalyzing the $\alpha_{IIb}\beta_3$ Solution Data

To obtain $< s^0_{(20,w)} >_w$ and $< D^0_{t(20,w)} >_z$ for the resting $\alpha_{IIb}\beta_3$ solubilized in OG, data derived from AUC and DLS runs, spanning more than five years of experiments on several batches (Hantgan et al., 1999, 2001, 2002, 2003), were plotted as $s^{-1}_{(20,w)}$ vs. c and $D_{t(20,w)}$ vs. c, respectively (Figure S2). As can be seen, while the individual errors are quite low ($s^{-1}_{(20,w)}$) or reasonable ($D_{t(20,w)}$), the values have a larger spread than usually found with "well-behaved" soluble proteins under near-physiological conditions. When all the points are fitted with a non-weighted straight line, $< s^0_{(20,w)} >_w = 8.04 \pm 0.11$ S and $< D^0_{t(20,w)} >_z = 2.95 \pm 0.15$ F are extrapolated, with SEM that are very good (1.4%) and satisfactory (5%), respectively (Table S1).

Since the experimental molecular weight (238,700, see Table S1) is in excellent agreement (2%) with that of $\alpha_{IIb}\beta_3$ (240,688, calculated from the primary sequence including carbohydrates and a 50-molecule OG micelle), the spread of the data suggests a conformationally heterogeneous population. To assess this hypothesis, the data were divided into two groups (indicated by the blue and magenta symbols in Figure S2), according to the values predicted by the global regression. By creating an additional variable g whose value will be 1 if the observed value is above the predicted value, and 0 if it is below, the data could be reanalyzed with the following equations:

$$s^{-1}_{(20,w)} = (s^0_{(20,w)})^{-1} + \left((s^0_{(20,w)})^{-1} k_s c\right) + \beta g \qquad (1)$$

$$D_{t(20,w)} = D^0_{t(20,w)} + \left(D^0_{t(20,w)} k_D c\right) + \gamma g \qquad (2)$$

where β and γ are the group-specific additional partial regression coefficients. It can be easily seen that equations (1) and (2) can each yield two intercepts, depending on the value of g. When $g = 0$, the intercepts will directly give the $(s^0_{(20,w)})^{-1}$ (and $D^0_{t(20,w)}$) values for that group, while for $g = 1$, the partial regression coefficients β (and γ) must be added to the respective intercepts to obtain the $(s^0_{(20,w)})^{-1}$ (and $D^0_{t(20,w)}$) values for the other group. This analysis was performed with Stata Statistical Software v.8.0 (Stata Corporation, College Station, TX). To assess if the two intercepts are statistically different, the null hypothesis $\beta = 0$ (or $\gamma = 0$) can be evaluated using Student's t test.

As reported in Table S1, the $< s^0_{(20,w)} >_w$ and $< D^0_{t(20,w)} >_z$ values extrapolated from the series with the "high" s and D values (■) are only ~2% and 4% higher, respectively, than the global regression values. The resulting molecular weight, somewhat lower but still within 3% of the calculated value, confirms the presence of a compact monomeric integrin species.

As for the "low" *s* and *D* series (■), the derived molecular weight is only ~7% higher than the calculated value (Table S1). These data are thus compatible with a more elongated but still mostly monomeric integrin species. Strikingly, the percent changes in *s* and *D* between the low- and the high-series values, -4.6% and -14%, respectively, are within the range of those induced by the peptide and non-peptide ligand-mimetics RGDF (-6.5% and –14%; Hantgan et al., 1999), eptifibatide (-5.7% and -10%; Hantgan et al., 2001), and tirofiban (-2.2% and –2.5%; Hantgan et al., 2002), that we previously reported.

The possibility that the spread in the *s* and *D* values is due to residual detergent used in the integrin purification, altering sample viscosity and density, is unlikely. Instead, these data suggest that each integrin batch has a variable proportion of at least two stable conformers, too close in frictional coefficients to be resolved by DLS or AUC. To test this hypothesis, Claverie simulations (Cohen and Claverie, 1975) were performed using the DCDT+ (version 6.31) software (J. Philo, Thousand Oaks, CA). The limiting values of $< s^0_{(20,w)} >_w$ (7.80 and 8.14 S; Table S1) were employed, with $< D^0_{t(20,w)} >_z$ values adjusted to give a constant MW of 240,000. Relative proportions of 20:80, 50:50, and 80:20 were examined. In every case, the simulations produced nearly symmetrical peaks that could be well fitted with a single species model (not shown), neatly recovering the input distribution, while the two-species fit worked only for the 50:50 case.

In conclusion, the high-series-derived $< s^0_{(20,w)} >_w$ and $< D^0_{t(20,w)} >_z$ values appear to be the best set against which to refine the resting integrin models. Similarly, the low-series-derived values will be used in modeling the primed conformation, but we will rely mainly on the $< s^0_{(20,w)} >_w$ value because of the greater sensitivity of $< D^0_{t(20,w)} >_z$ to aggregates.

Supplemental References

Cohen, R., and Claverie, J.M. (1975). Sedimentation of generalized systems of interacting particles. II. Active enzyme centrifugation--theory and extensions of its validity range. Biopolymers *14*, 1701-1716.

Mehta, R.J., Diefenbach, B., Brown, A., Cullen, E., Jonczyk, A., Güssow, D., Luckenbach, G.A., and Goodman, S.L. (1998). Transmembrane-truncated αvβ3 integrin retains high affinity for ligand binding: evidence for an ' inside-out ' suppressor. Biochem. J. *330*, 861-869.

Rocco, M., Spotorno, B, and Hantgan, R.R. (1993). Modeling the $\alpha_{IIb}\beta_3$ integrin solution conformation. Protein Sci. *2*, 2154-2166.

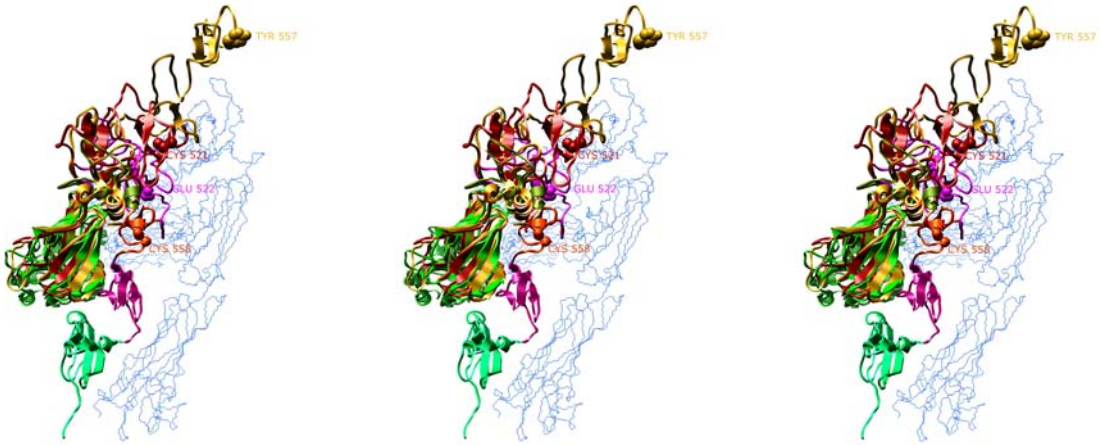

Figure S1. Stereo Views of the PSI/Hybrid/I-EGF1-3 β_2 Homology Models Superimposed on the Crystal Structure-Based Av-3 Model of the $\alpha_v\beta_3$ Ectodomain

In this image, the left-center stereo pair can be viewed in crossed-eyes mode, while the center-right pair is for relaxed eyes mode. The α_v chain in model Av-3 is shown as a light blue trace, while the β_3 chain is shown as ribbons color-coded as in Fig. 1, panel A (main text). The homology models based on the β_2 PSI/hybrid/I-EGF1-2 (2P26) and PSI/hybrid/I-EGF1-3 (2P28) crystal structures (Shi et al., 2007) are colored firebrick red and gold, respectively, and are superimposed on the Av-3 model using the hybrid domain as a template. The C-terminal ends Cys521 and Tyr557 of the two β_2 homology models are shown in space-filling mode, as well as the corresponding residues (Glu522 and Cys558) in the Av-3 model where they should be respectively attached to restore the chain connectivity.

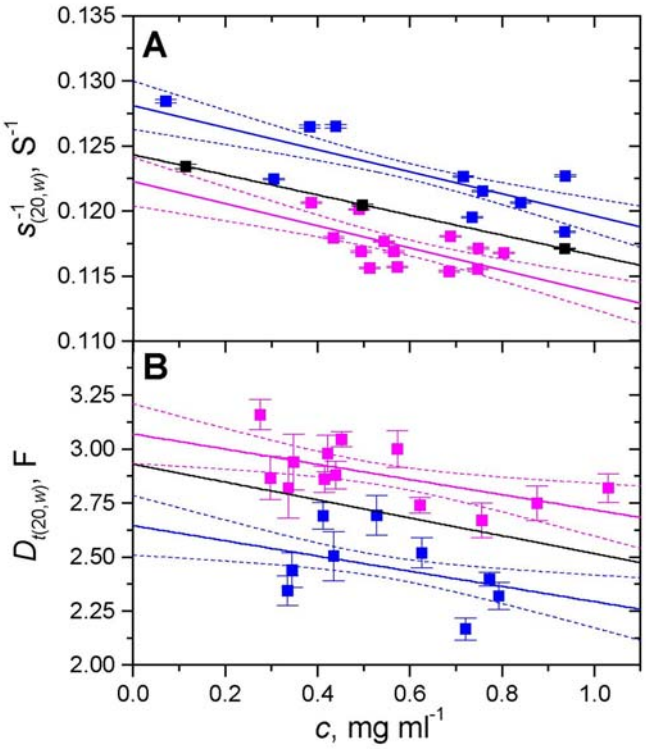

Figure S2. Plots and Analysis of the $s^{-1}_{(20,w)}$ and $D_{t(20,w)}$ Experimental Data

Panel A: $s^{-1}_{(20,w)}$ vs. c. Panel B: $D_{t(20,w)}$ vs. c. The black lines represent the global linear regressions for each set, according to the equations $s^{-1}_{(20,w)} = (s^{0}_{(20,w)})^{-1}\left(1 + k_s c\right)$ and $D_{t(20,w)} = D^{0}_{t(20,w)}\left(1 + k_D c\right)$. Points lying above and below the black lines in both panels were then assigned to two groups (■ and ■, respectively; ■, not assigned) and globally re-fitted with equations (1) and (2), respectively (see Supplemental Results). The solid blue and magenta lines are the resulting linear predictions for the two groups, with their 90% confidence intervals (dashed lines).

Table S1. A Summary of the Hydrodynamic Parameters and Related Properties (± SEM) Derived from the Linear Regressions of the Experimental Data Reported in Figure S2

Data included	$< s^0_{(20,w)} >_w$ (S)	r^2 (*P*)	k_S (ml g^{-1})	$< D^0_{t(20,w)} >_z$ (F)	r^2 (*P*)	$K_D \times 10^{-6}$ (ml g^{-1})	$<M>_{w/z}$,[a] (g mol^{-1})
All	8.04 ± 0.11	0.2427	-62.2 ± 22.5	2.95 ± 0.15	0.1092	-1.44 ± 0.90	238,700 ± 12,600
High set (■)	8.18 ± 0.07	0.8156	-69.4 ± 13.9	3.07 ± 0.08	0.7645	-1.33 ± 0.53	233,400 ± 6,400
Low set (■)	7.80 ± 0.08	(<0.001)[b]	-66.2 ± 13.3	2.65 ± 0.09	(<0.001)[c]	-1.15 ± 0.46	257,800 ± 7,100

[a]Computed from the $< s^0_{(20,w)} >_w$ and $< D^0_{t(20,w)} >_z$ values via the Svedberg equation, using a partial specific volume $\bar{v}_2$ = 0.723 ml g^{-1};

[b]$t_{(23)}$ = (β/SEM(β)) = (0.00586/0.00071) = 8.23;

[c]$t_{(19)}$ = (γ/SEM(γ)) = (0.425/0.058) = 7.27.