

## Combined Experimental and Theoretical Study of Long-Range Interactions Modulating Dimerization and Activity of Yeast Geranylgeranyl Diphosphate Synthase

Chia-Hsiang Lo,<sup>†</sup> Ying-Hsuan Chang,<sup>†</sup> Jon D. Wright,<sup>§</sup> Shih-Hsun Chen,<sup>||</sup>  
Daphne Kan,<sup>‡</sup> Carmay Lim,<sup>\*,§</sup> and Po-Huang Liang<sup>\*,†,‡</sup>

*Institute of Biochemical Sciences, National Taiwan University, Taipei; Institute of Biological Chemistry and Biomedical Sciences, Academia Sinica, Taipei; and Department of Biological Science and Technology, National Chiao Tung University, Hsin-Chu, Taiwan*

Received November 6, 2008; E-mail: phliang@gate.sinica.edu.tw; carmay@gate.sinica.edu.tw

**Abstract:** We present here how two amino acid residues in the first helix distal from the main dimer interface modulate the dimerization and activity of a geranylgeranyl diphosphate synthase (GGPPs). The enzyme catalyzes condensation of farnesyl diphosphate and isopentenyl diphosphate to generate a C<sub>20</sub> product as a precursor for chlorophylls, carotenoids, and geranylgeranylated proteins. The 3D structure of GGPPs from *Saccharomyces cerevisiae* reveals an unique positioning of the N-terminal helix A, which protrudes into the other subunit and stabilizes dimerization, although it is far from the main dimer interface. Through a series of mutants that were characterized by analytic ultracentrifugation (AUC), the replacement of L8 and I9 at this helix with Gly was found sufficient to disrupt the dimer into a monomer, leading to at least 10<sup>3</sup>-fold reduction in activity. Molecular dynamics simulations and free energy decomposition analyses revealed the possible effects of the mutations on the protein structures and several critical interactions for maintaining dimerization. Further site-directed mutagenesis and AUC studies elucidated the molecular mechanism for modulating dimerization and activity by long-range interactions.

### Introduction

There are growing numbers of examples in which dimerization is required for enzyme activity. For example, the protease from human immuno-deficiency virus (HIV) has an active site formed by two monomers, each of which provides a catalytic Asp residue. The dimer interface is a potential target for an anti-HIV drug.<sup>1</sup> The protease from severe acute respiratory syndrome coronavirus<sup>2</sup> and the protease from human cytomegalovirus show no enzyme activity in the monomeric state.<sup>3,4</sup> A dipeptide prolyl protease 4, which cleaves the peptide bond after the penultimate residue, requires a conserved His at the dimer interface for dimer stability; mutation of this His to Glu was found to disrupt dimerization and to decrease enzyme activity by 300-fold.<sup>5</sup> Whereas previous work showed that the mutation of a residue in the dimer interface disrupted dimerization, this work reveals mutation of two amino acid (aa) residues at the N-terminus of an enzyme, geranylgeranyl diphosphate synthase (GGPPs), distal from the main dimer interface can also disrupt dimerization and decrease enzymatic

activity. Further insight as to how mutations distal from the main dimer interface can modulate dimerization was obtained from molecular dynamics (MD) simulations of the wild-type and mutant GGPPs and subsequent free energy decomposition analyses.

GGPPs catalyzes the head-to-tail condensation of C<sub>15</sub> farnesyl diphosphate (FPP) with a C<sub>5</sub> isopentenyl diphosphate (IPP) to form the C<sub>20</sub> geranylgeranyl diphosphate (GGPP) product.<sup>6–8</sup> This compound can be used to make chlorophylls or  $\alpha$ -tocopherol,<sup>9</sup> ent-kaurene, or taxadiene.<sup>10</sup> It can be further elongated to produce long-chain isoprenoid used in quinine biosynthesis.<sup>9</sup> Furthermore, two GGPP molecules can condense to form phytoene,<sup>11</sup> the precursor for many carotenoids. GGPP or FPP can be used as a ligand for protein prenylation, an essential post-translational modification for signaling proteins such as Ras, Rho, Rab, and Rac.<sup>12,13</sup> Thus, FPP synthase (FPPs), which

<sup>†</sup> National Taiwan University.

<sup>‡</sup> Institute of Biological Chemistry, Academia Sinica.

<sup>§</sup> Institute of Biomedical Sciences, Academia Sinica.

<sup>||</sup> National Chiao Tung University.

(1) De Clercq, E. *Nat. Rev. Drug Discovery* **2002**, *1*, 13–25.

(2) Fan, K.; Wei, P.; Feng, Q.; Chen, S.; Huang, C.; Ma, L.; Lai, B.; Pei, J.; Liu, Y.; Chen, J.; Lai, L. *J. Biol. Chem.* **2004**, *279*, 1637–1642.

(3) Drake, P. L.; Cole, J. L.; Waxman, L.; Hall, D. L.; Sardana, M. K.; Kuo, L. C. *J. Biol. Chem.* **1996**, *271*, 7445–7449.

(4) Margosiak, S. A.; Vanderpool, D. L.; Sisson, W.; Pinko, C.; Kan, C. C. *Biochemistry* **1996**, *35*, 5300–5307.

(5) Chien, C. H.; Huang, L. H.; Chou, C. Y.; Chen, Y. S.; Han, Y. S.; Chang, G. G.; Liang, P. H.; Chen, X. *J. Biol. Chem.* **2004**, *279*, 52338–52345.

(6) Sacchettini, J. C.; Poulter, C. D. *Science* **1997**, *277*, 1788–1789.

(7) Ogura, K.; Koyama, T. *Chem. Rev.* **1998**, *98*, 1263–1276.

(8) Liang, P. H.; Ko, T. P.; Wang, A. H. *Eur. J. Biochem.* **2002**, *269*, 3339–3354.

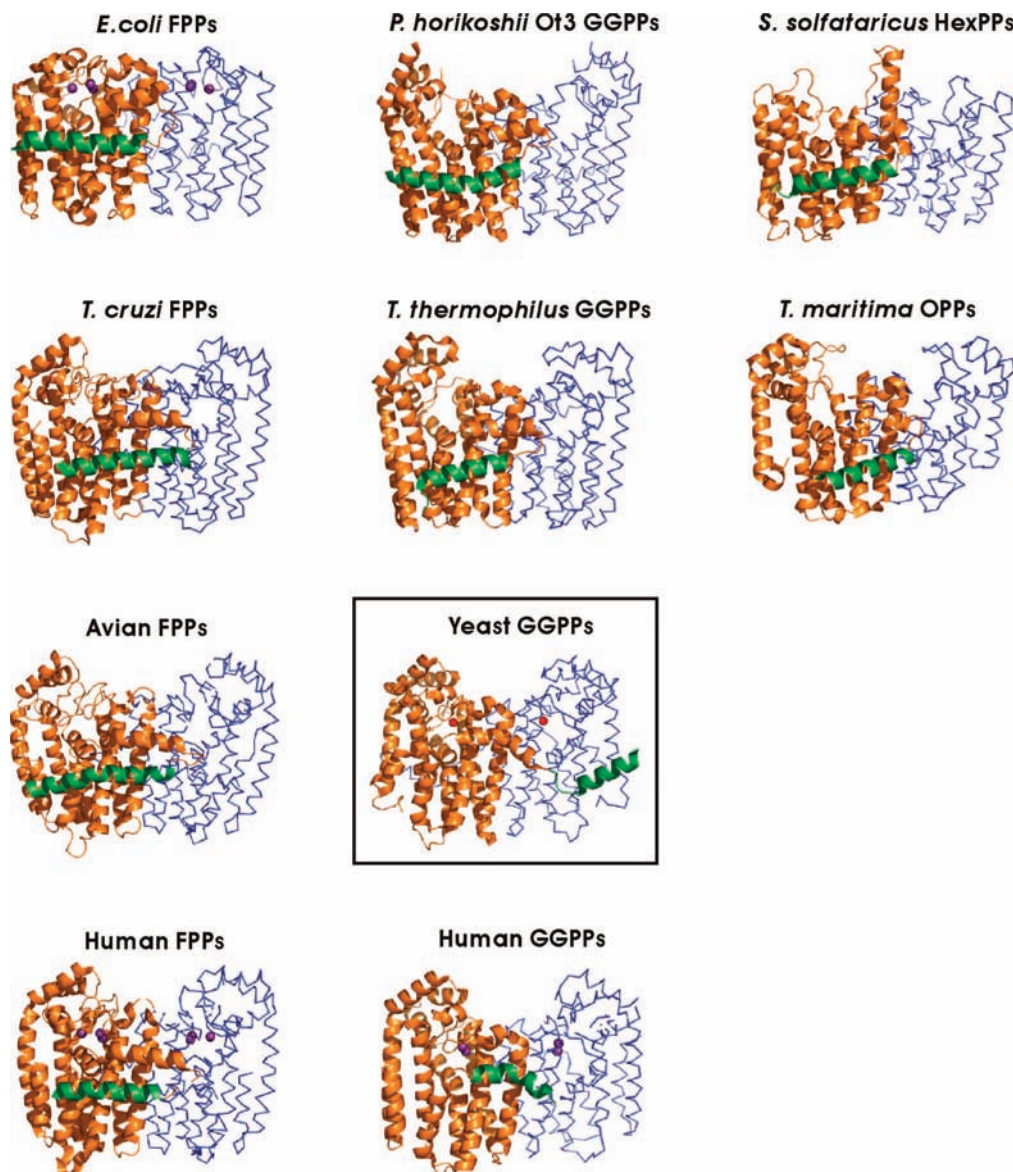
(9) Rüdiger, W.; Benz, J.; Guthoff, C. *Eur. J. Biochem.* **1980**, *109*, 193–200.

(10) Jiang, Y.; Proteau, P.; Poulter, D.; Ferro-Novick, S. *J. Biol. Chem.* **1995**, *270*, 21793–21799.

(11) Olszewski, N.; Sun, T.; Gubler, F. *Plant Cell* **2002**, *14*, S61–S80.

(12) Dogbo, O.; Laferrriere, A.; D'Harlingue, A.; Camara, B. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 7054–7058.

(13) Kuzuguchi, T.; Morita, Y.; Sagami, I.; Sagami, H.; Ogura, K. *J. Biol. Chem.* **1999**, *274*, 5888–5894.



**Figure 1.** Ribbon representation of the dimer structures of *trans*-prenyltransferases. From left to right are the structures of FPPs in the 1st column (PDB entries 1RQI, 1YHL, 1FPS, and 1YV5, for *Escherichia coli*, *Trypanosoma cruzi*, avian, and human FPPs, respectively), GGPPs in the 2nd column (PDB entries 1WY0, 1WMW, 2DH4, and 2FVI for *Pyrococcus horikoshii* Ot3, *Thermus thermophilus*, yeast, and human GGPPs, respectively), HexPPs and OPPs in the 3rd column (PDB entries 2AZJ and 1V4E for *Sulfolobus solfataricus* HexPPs and *Thermotoga maritima* OPPs, respectively). In these structures, the left subunit is displayed in orange with the first helix in green, and the right subunit is in blue. Only in the yeast GGPPs structure (boxed), the first helix is protruded from its own subunit to bind with the opposite subunit. The helix A is distant from the main dimer interface formed by residues mainly from helices F and G. Figures plotted using the PyMol program.<sup>38</sup>

catalyzes FPP through condensation of C<sub>10</sub> geranyl diphosphate with IPP, has been used as a drug target to develop bisphosphonate drugs for treatment of bone resorption diseases such as osteoporosis and infection by parasitic protozoa.<sup>14</sup> One of the drugs was thought to target GGPPs.<sup>15</sup>

GGPPs belongs to the *trans*-prenyltransferase family, since a *trans*-double bond is formed during the IPP condensation with FPP. The members in this family, which synthesize the final products of C<sub>15</sub> by FPPs, C<sub>20</sub> by GGPPs, C<sub>30</sub> by hexaprenyl

diphosphate synthase (HexPPs), and C<sub>40</sub> by octaprenyl diphosphate synthase (OPPs), share sequence homology and possess similar 3D structures composed of 15  $\alpha$ -helices connected by loops.<sup>16–22</sup> However, the GGPPs from *Saccharomyces cerevisiae* shows a unique positioning of the first N-terminal helix (see Figure 1), which protrudes into and binds the opposite subunit.<sup>20</sup> Deletion of this helix led to re-establishment of an

(14) Guo, R. T.; Cao, R.; Liang, P. H.; Ko, T. P.; Chang, T. H.; Hudock, M. P.; Jeng, W. Y.; Chen, C.; Zhang, Y.; Song, Y.; Kuo, C. J.; Yin, F.; Oldfield, E.; Wang, A. H. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 10022–10027.

(15) Goffinet, M.; Thoulouzan, M.; Pradines, A.; Lajoie-Mazenc, I.; Weinbaum, C.; Faye, J. C.; Seronie-Vivien, S. *BMC Cancer* **2006**, *6*, 60.

(16) Guo, R. T.; Kuo, C. J.; Chou, C. C.; Ko, T. P.; Shr, H. L.; Liang, P. H.; Wang, A. H. *J. Biol. Chem.* **2004**, *279*, 4903–4912.

(17) Hosfield, D. J.; Zhang, Y.; Dougan, D. R.; Broun, A.; Tari, L. W.; Swanson, R. V.; Finn, J. *J. Biol. Chem.* **2004**, *279*, 8526–8529.

(18) Sun, H. Y.; Ko, T. P.; Kuo, C. J.; Guo, R. T.; Chou, C. C.; Liang, P. H.; Wang, A. H. *J. Bacteriol.* **2005**, *187*, 8137–8148.

(19) Gabelli, S. B.; McLellan, J. S.; Montalvetti, A.; Oldfield, E.; Docampo, R.; Amzel, L. M. *Proteins* **2006**, *62*, 80–88.

(20) Chang, T. H.; Guo, R. T.; Ko, T. P.; Wang, A. H. J.; Liang, P. H. *J. Biol. Chem.* **2006**, *281*, 14991–15000.

equilibrium between the dimeric and monomeric forms of the enzyme, but greatly favored monomer, and significant loss of enzymatic activity.<sup>20</sup>

As helix A is distant from the main dimer interface (~25 Å), it was not understood (i) how it contributes to the GGPPs subunit dimerization and (ii) whether the entire helix is required for dimerization and enzyme activity. In this study, site-directed mutagenesis, in conjunction with analytical ultracentrifugation (AUC) measurements and molecular dynamics (MD) simulations, was used to gain insight into the long-range interactions modulating dimerization.

## Materials and Methods

**Materials.** *PfuTurbo*, the plasmid miniprep kit, DNAGel extraction kit, and Ni-NTA resin were purchased from Qiagen. The protein expression kit (including the pET32Xa/LIC vector and competent JM109 and BL21 cells) was obtained from Novagen. The QuikChange site-directed mutagenesis kit was obtained from Stratagene. Radiolabeled [<sup>14</sup>C]IPP (55 mCi/mmol) was purchased from Amersham Pharmacia Biotech. Nonlabeled FPP and IPP were obtained from Sigma. All commercial buffers and reagents were of the highest grade.

**Plasmid Construction for Mutant GGPPs.** The gene encoding GGPPs was cloned from *S. cerevisiae* genomic DNA as previously reported.<sup>20</sup> GGPPs mutants were prepared by using the QuikChange site-directed mutagenesis kit. The mutagenic oligonucleotides are 5'-atggaggccaagatagatGGGctgatcaataatgatcctgttgg-3' and 5'-ccaaacagatcattatgatCCcctatctcttggcctccat-3' for E7G, 5'-atggaggccaagatagatgagGGGatcaataatgatcctgttgg-3' and 5'-ccaaacagatcattatgatCCCctcatctcttggcctccat-3' for L8G, 5'-atggaggccaagatagatgagctGGGaaataatgatcctgttgg-3' and 5'-ccaaacagatcattatCCCcagctcatctcttggcctccat-3' for I9G, 5'-atggaggccaagatagagGGGGGatcaataatgatcctgttgg-3' and 5'-ccaaacagatcattatgatCCCCctcatctcttggcctccat-3' for E7G/L8G, 5'-atggaggccaagatagatgagGGGGGaaataatgatcctgttgg-3' and 5'-ccaaacagatcattatCCCCctcatctcttggcctccat-3' for L8G/I9G, and 5'-ggtccagcaaaatgaaagcGGGGTcaaacacctat-3' and 5'-ataagtttgaACCCCCgttcttattgctgctgac-3' for L22G/I23G; the mutagenic oligonucleotides for performing truncated mutagenesis are 5'-ggtattgagggtcgcgagctgatacaataatgatcctgt-3' and 5'-agaggagattagagctcacaattcggataagtgctc-3' for Δ(1–6), 5'-ggtattgagggtcgcctgatacaataatgatcctgttgg-3' and 5'-agaggagattagagctcacaattcggataagtgctc-3' for Δ(1–7), and 5'-ggtattgagggtcgcctgatacaataatgatcctgttgg-3' and 5'-agaggagattagagctcacaattcggataagtgctc-3' for Δ(1–8), Δ(1–9) and Δ(1–17) were previously generated.<sup>21</sup>

The mutagenic oligonucleotides for performing site-directed mutagenesis to probe the important interactions in stabilizing dimerization are 5'-ggagatgatGGGaataggtGGGaaataaacagcggg-3' and 5'-ccgctgttttatiCCCaacatattCCCatacatccc-3' for L163G/M167G, 5'-ggttcctttataaatcttGGGggtattGGGtatcagattagatg-3' and 5'-catctctaattgataCCCaataccCCCaaattataaaaggaacc-3' for L200G/I203G, 5'-ggagatgatGGGaataggttattgaataaacagcggg-3' and 5'-ccgctgttttatacaaacatattCCCatacatccc-3' for L163G, 5'-ggagatgattgaataggttGGGaaataaacagcggg-3' and 5'-ccgctgttttattCCCaacatattcaaacatccc for M167G, 5'-ggttcctttataaatcttGGGggtatttattatcagattagatg-3' and catctctaattgataaaataaacCCCaaattataaaaggaacc-3' for L200G, 5'-ggttcctttataaatcttgggtattGGGtatcagattagatg-3' and 5'-catctctaattgataCCCaataccaaaagattataaaaggaacc-3' for I203G, 5'-ggtgtaccctccactataGGCaccgaaattatag-3' and 5'-catataatttgcggtGGCtatatggagggtacacc-3' for N101G, 5'-ggacaaggcttAAGatatactggagagacttctgcc-3' and 5'-ggcagaagctctcagatataCTTcaagcctgtcc-3' for D145K, 5'-gattacgattttcaacgaaGCAttgatcaatctacataggggacaag-3' and 5'-ctgtccctatgatattgatca-

TGCTctgtgaaatcgtaatc-3' for E134A, 5'-aaaacagcggccttttGCAtaacgttgagactatg-3' and 5'-catgagctcaacgttaaTGCgaaagggcgcctgttt-3' for R175A, and 5'-ggcattcgttggctctttataGCTcttctgggtattatcatcag-3' and 5'-ctgataataataccagaagAGCtataaaaggaaccaacgaatggcc-3' for M167G/N199A (with M167G as template).

**Expression and Purification of the Mutant GGPPs.** The mutant GGPPs plasmids were used to transform *E. coli* JM109 competent cells that were streaked on a Luria–Bertanini (LB) agar plate containing 100 μg/mL ampicillin. Ampicillin-resistant colonies were selected from the agar plate and grown in 5 mL of LB culture containing 100 μg/mL ampicillin overnight at 37 °C. The correct constructs confirmed by sequencing were transformed to *E. coli* BL21(DE3) for protein expression. The 60 mL of overnight culture of a single transformant was used to inoculate 6 L of fresh LB medium containing 100 μg/mL ampicillin. The cells were grown to  $A_{600} = 0.6$  and induced with 1 mM IPTG at 16 °C. After 16 h, the cells were harvested by centrifugation at 7000g for 15 min to collect the cell paste. The enzyme purification was conducted at 4 °C. Cell paste was suspended in 75 mL of lysis buffer containing 25 mM Tris-HCl, at pH 7.5, and 150 mM NaCl. Cell lysate was prepared with a French pressure cell press (AIM-AMINCO Spectronic Instruments) and centrifuged at 17 000g to remove cell debris. The cell-free extract was loaded onto the Ni-NTA column, which had been previously equilibrated with lysis buffer. The column was washed with 10 mM imidazole followed by 20 mM imidazole-containing buffer. His-tagged GGPPs eluted with 100 mM imidazole was dialyzed twice against 3 L of buffer (25 mM Tris-HCl, pH 7.5, and 150 mM NaCl) and then subjected to the Factor Xa (FXa) protease digestion to remove the tag. The mixture was then passed through another Ni-NTA column, and subsequently, untagged GGPPs was eluted with 10 mM imidazole-containing buffer and then dialyzed twice against 3 L of buffer (25 mM Tris-HCl, pH 7.5, and 150 mM NaCl) for storage. SDS-PAGE analysis was used to check the purity of GGPPs and its mutants.

**AUC Experiments.** AUC measurements of the purified mutant GGPPs proteins were performed at the approximate concentrations of 0.25–1.0 mg/mL. The sedimentation coefficients (*S*) of the enzymes were estimated with a Beckman–Coulter XL-A analytical ultracentrifuge with an An60Ti rotor. The sedimentation velocity experiment was performed at 40 000 rpm at 20 °C with a standard double-sector aluminum centerpiece. Both sets of data were analyzed with the program Sedfit version 9.4c (<http://www.analyticalultracentrifugation.com>) to calculate the molecular weights and sedimentation coefficients as previously described.<sup>23</sup> The Sednterp version 1.09 program (<http://www.jphilo.mailway.com/>) was used to calculate solvent densities and viscosities. The observed sedimentation profiles of a continuous size distribution (*c(s)*) can be calculated from the following equation:  $a(r, t) = \int f(c(s))L(s, D, r, t)ds + \epsilon$ , where  $a(r, t)$  denotes the experimentally observed signal.  $L(s, D, r, t)$  denotes the solution of the Lamm equation for a single species and  $\epsilon$  is the noise component. The dissociation constant ( $K_d$ ) of the dimer–monomer equilibrium is calculated by global fitting of the sedimentation data with the SEDPHAT program.

**Kinetic Measurements of the Mutant GGPPs.** For enzymatic activity measurements, each mutant GGPPs (0.2 μM E7G, L8G, I9G, E7G/L8G, L8G/I9G, Δ(1–6), Δ(1–7), Δ(1–8), Δ(1–9), or Δ(1–17)) was used. The reaction was initiated in 200 μL of solution containing 100 mM Hepes (pH 7.5), various concentrations of FPP and [<sup>14</sup>C]IPP as specified below, 50 mM KCl, 0.5 mM MgCl<sub>2</sub>, and 0.1% Triton X-100 at 25 °C. The enzyme concentration used in all experiments was determined from its absorbance at 280 nm ( $\epsilon = 20\,340\text{ M}^{-1}\text{cm}^{-1}$ ). Measurements of the kinetic parameters for the wild-type and the mutants followed our published procedure.<sup>24</sup> For the IPP  $K_m$  determinations, 25 μM FPP was utilized to saturate the enzyme, and the IPP concentrations from 0.25–50 μM varied with the  $K_m$  of IPP for each mutant were employed. For the FPP  $K_m$  and  $k_{cat}$  measurements, 0.25–50 μM FPP was used along with 50

(21) Kavanagh, K. L.; Dunford, J. E.; Bunkoczi, G.; Russell, R. G. G.; Oppermann, U. *J. Biol. Chem.* **2006**, *281*, 22004–22012.

(22) Kloer, D. P.; Welsch, P.; Beyer, P.; Schulz, G. E. *Biochemistry* **2006**, *45*, 15197–15204.

(23) Chang, H. C.; Chang, G. G. *J. Biol. Chem.* **2004**, *278*, 23996–24002.

$\mu\text{M}$  [ $^{14}\text{C}$ ]IPP. To measure the initial rate, 40- $\mu\text{L}$  portions of the reaction mixture were periodically withdrawn within 10% substrate depletion and then mixed with 10 mM EDTA for reaction termination. The radiolabeled products were then extracted with 1-butanol, and the radioactivity associated with aqueous and butanol phases were separately quantified by using a Beckmann LS6500 scintillation counter. Data of initial rates versus substrate concentrations were analyzed by nonlinear regression of the Michaelis–Menten equation using the KaleidaGraph (Synergy software) to obtain the  $K_m$  and  $V_{\text{max}}$  values.<sup>25</sup> The  $k_{\text{cat}}$  was calculated from  $V_{\text{max}}/[\text{E}]$ .

**3D Model Building of Wild-Type GGPPs.** The coordinates of the GGPPs dimer (aa –3 to 335), including the bound  $\text{Mg}^{2+}$  for each subunit, were taken from the 1.98-Å crystal structure, PDB entry 2DH4 (the first five residues, numbered –4 to 0, is a linker to expose the FXa cleavage site, but the coordinates of the N-terminal residue, Met –4, were missing). For each chain, the missing loop between residues 310 and 322 was modeled using the MODLOOP program<sup>26,27</sup> and inserted into the 3D structure. Hydrogen atoms were added and minimized using the CHARMM<sup>28</sup> version 34 program with the CHARMM27 all-atom parameter set.<sup>29</sup> At the crystallization pH of 7.5, all Asp and Glu residues were deprotonated, Arg and Lys residues were protonated, while His residues were protonated according to the local environment using the Whatif program.<sup>30</sup> This resulted in a net dimer charge of  $-12e$ , which was neutralized by adding 12 sodium counterions at the highest electronegativity locations with the constraints that each counterion was  $\geq 5$  Å from the protein surface and  $\geq 9$  Å from each other. The resultant system was solvated in a truncated octahedron containing TIP3P water molecules,<sup>31</sup> resulting in a total of 145 617 atoms.

**MD Simulations of Wild-Type and Mutant GGPPs.** To relieve any bad contacts in the solvated wild-type model structure, the water molecules were subjected to rounds of minimization with constraints on the protein heavy atoms. The resulting solvated system was subjected to MD at a mean temperature of 300 K using a 1 fs time step, periodic boundary conditions, van der Waals (vdW) interactions shifted to zero at 12 Å, and electrostatic interactions treated via the particle mesh Ewald summation method.<sup>32</sup> Initially, 50 ps of dynamics was performed with the protein atoms and  $\text{Mg}^{2+}$  ions restrained by a harmonic potential, which was then reduced for all protein atoms and removed for the  $\text{Mg}^{2+}$  ions after 50 ps. This was followed by another 200 ps dynamics for the modeled loop region (residues 310–322). The final structure (without water molecules) was used as the starting point for simulations of the wild-type GGPPs dimer and the L8G/I9G double mutant dimer (by mutating the L8 and I9 side-chains to Gly). The root-mean-square deviation (rmsd) of the protein backbone in this starting structure from that in the X-ray structure is  $\sim 0.5$  Å.

Each structure was solvated again in a truncated octahedron containing TIP3P water molecules. Subsequently, 50 ps of dynamics was performed allowing only water molecules to move, followed by 50 ps with the protein backbone atoms lightly restrained, and

another 50 ps allowing all atoms in loop regions to move, and finally 900 ps of unconstrained dynamics. The backbone rmsd from the initial structure is  $2.1 \pm 0.1$  Å for the wild-type dimer and  $2.9 \pm 0.1$  Å for the double mutant dimer. Coordinates were saved every 0.1 ps from the final 100 ps of dynamics for a total of 1000 coordinate sets for each GGPPs structure. These were used to compute a mean MD structure and used in the free-energy decomposition analyses below.

**Free Energy Decomposition.** The dimerization free energy of the wild-type or mutant protein (denoted by p) in solution was based on the thermodynamic cycle: It is given by

$$\begin{array}{ccc} \Delta G_{\text{gas}}(\text{p}) & & \\ 2[\text{p}]_{\text{gas}} & \rightarrow & [\text{p}\cdot\text{p}]_{\text{gas}} \\ -2G_{\text{solv}}(\text{p}) \uparrow & & \downarrow G_{\text{solv}}(\text{p}\cdot\text{p}) \\ 2[\text{p}]_{\text{sln}} & \rightarrow & [\text{p}\cdot\text{p}]_{\text{sln}} \\ \Delta G_{\text{sln}}(\text{p}) & & \\ \Delta G_{\text{sln}}(\text{p}) = \Delta G_{\text{gas}}(\text{p}) + G_{\text{solv}}(\text{p}\cdot\text{p}) - 2G_{\text{solv}}(\text{p}) & (1) & \end{array}$$

As conformational changes upon dimerization were neglected (see below), the gas-phase dimerization free energy,  $\Delta G_{\text{gas}}$ , was approximated as a sum of the changes in the gas-phase vdW ( $\Delta E_{\text{gas}}^{\text{vdW}}$ ) and electrostatic ( $\Delta E_{\text{gas}}^{\text{elec}}$ ) energies upon dimerization, which were calculated using the CHARMM program and forcefield with a dielectric constant of 1 and a nonbond cut-off of 999 Å. The solvation free energy,  $G_{\text{solv}}$ , was estimated as a sum of the nonelectrostatic ( $G_{\text{solv}}^{\text{nonel}}$ ) and electrostatic ( $G_{\text{solv}}^{\text{elec}}$ ) contributions. The  $G_{\text{solv}}^{\text{nonel}}$  was approximated by a linear function of the solvent accessible surface area (SASA); that is,  $G_{\text{solv}}^{\text{nonel}} = \gamma \times \text{SASA}$ , where  $\gamma = 7.2 \text{ cal/mol}/\text{Å}^2$ ,<sup>33,34</sup> and the absolute SASA of the protein was computed using the CHARMM program and forcefield with a solvent probe radius of 1.4 Å. The  $G_{\text{solv}}^{\text{elec}}$  was estimated by the finite-difference solution to the linearized Poisson–Boltzmann equation implemented in the CHARMM program, using a set of radii optimized for solvation calculations using the CHARMM forcefield.<sup>35</sup> The contribution of an individual residue  $i$  to the dimerization free energy can be determined from (a)  $\Delta E_{\text{gas}}^{\text{vdW}}(i)$  and  $\Delta E_{\text{gas}}^{\text{elec}}(i)$ , the pairwise vdW and electrostatic interactions of residue  $i$  in chain A with all residues in chain B, respectively, (b)  $G_{\text{solv}}^{\text{nonel}}(i) = \gamma \times \text{SASA}(i)$ , where  $\text{SASA}(i)$  is the SASA of residue  $i$  in the dimer/monomer, and (c)  $G_{\text{solv}}^{\text{elec}}(i)$ , which can be decomposed because, in using the linearized Poisson–Boltzmann equation, superposition allows the net electrostatic potential at any point to be expressed as a sum of contributions from the atomic charges of residue  $i$ . Thus,  $G_{\text{solv}}^{\text{elec}}(i)$  can be computed by summing over all charges, the product of the charge and the potential at the position of the charge due to the atomic charges from residue  $i$  in the dimer/monomer. We refer the reader to previous works<sup>36,37</sup> for details of the per-residue free energy calculations.

The 1000 coordinate sets saved during the final 100 ps of each simulation were used to compute the free energies in eq 1. For each coordinate set, the first residue (Thr –3), which was found to be very flexible, was omitted as well as all water molecules and counter-ions to eliminate issues with boundary sizes. Following previous work,<sup>37</sup> the 1000 coordinate sets were sorted into 10 evenly distributed clusters based on the  $\Delta E_{\text{gas}}^{\text{elec}}$  between the two GGPPs chains computed using a 18 Å cutoff for each coordinate set (see

(24) Pan, J. J.; Chiou, S. T.; Liang, P. H. *Biochemistry* **2000**, *39*, 10936–10942.

(25) Eftink, M. R. *Methods Enzymol.* **1995**, *259*, 478–512.

(26) Fiser, A.; Do, R. K. G.; Sali, A. *Protein Sci.* **2000**, *9*, 1753–1773.

(27) Fiser, A.; Sali, A. *Bioinformatics* **2003**, *18*, 2500–2501.

(28) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.

(29) MacKerell, J. A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(30) Vriend, G. *J. Mol. Graph.* **1990**, *8*, 52–56.

(31) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(32) Feller, S. E.; Pastor, R. W.; Rojnickarin, A.; Bogusz, S.; Brooks, B. R. *J. Phys. Chem.* **1996**, *100*, 17011–17020.

(33) Sharp, K. A.; Nicholls, A.; Fine, R. F.; Honig, B. *Science* **1991**, *252*, 106–109.

(34) Jayaram, B.; McConnell, K. J.; Surjit, B. D.; Beveridge, D. L. *J. Comp. Phys.* **1999**, *151*, 333–357.

(35) Nina, M.; Beglov, D.; Roux, B. *J. Phys. Chem. B* **1997**, *101*, 5239–5248.

**Table 1.** Hydrodynamic Properties of Wild-Type and Mutant GGPPs

yeast GGPPs	major quaternary structure	sedimentation coefficient (S)	molecular weight (kDa)	frictional ratio ( $f/f_0$ )
wild-type	dimer	4.8	78	1.19
$\Delta(1-6)$	dimer	4.8	75	1.20
$\Delta(1-7)^a$	mixture	N.A.	N.A.	1.14
$\Delta(1-8)$	monomer	3.5	38	1.06
$\Delta(1-9)$	monomer	3.5	38	0.99
$\Delta(1-17)$	monomer	3.5	37	1.02
E7G	dimer	4.8	78	1.18
L8G <sup>a</sup>	mixture	N.A.	N.A.	1.18
I9G	monomer	3.4	40	1.09
E7G/L8G	monomer	3.4	39	1.08
L8G/I9G	monomer	3.4	38	1.08

<sup>a</sup> For mixture of dimer and monomer, sedimentation coefficient and molecular weight cannot be determined (N.A.).

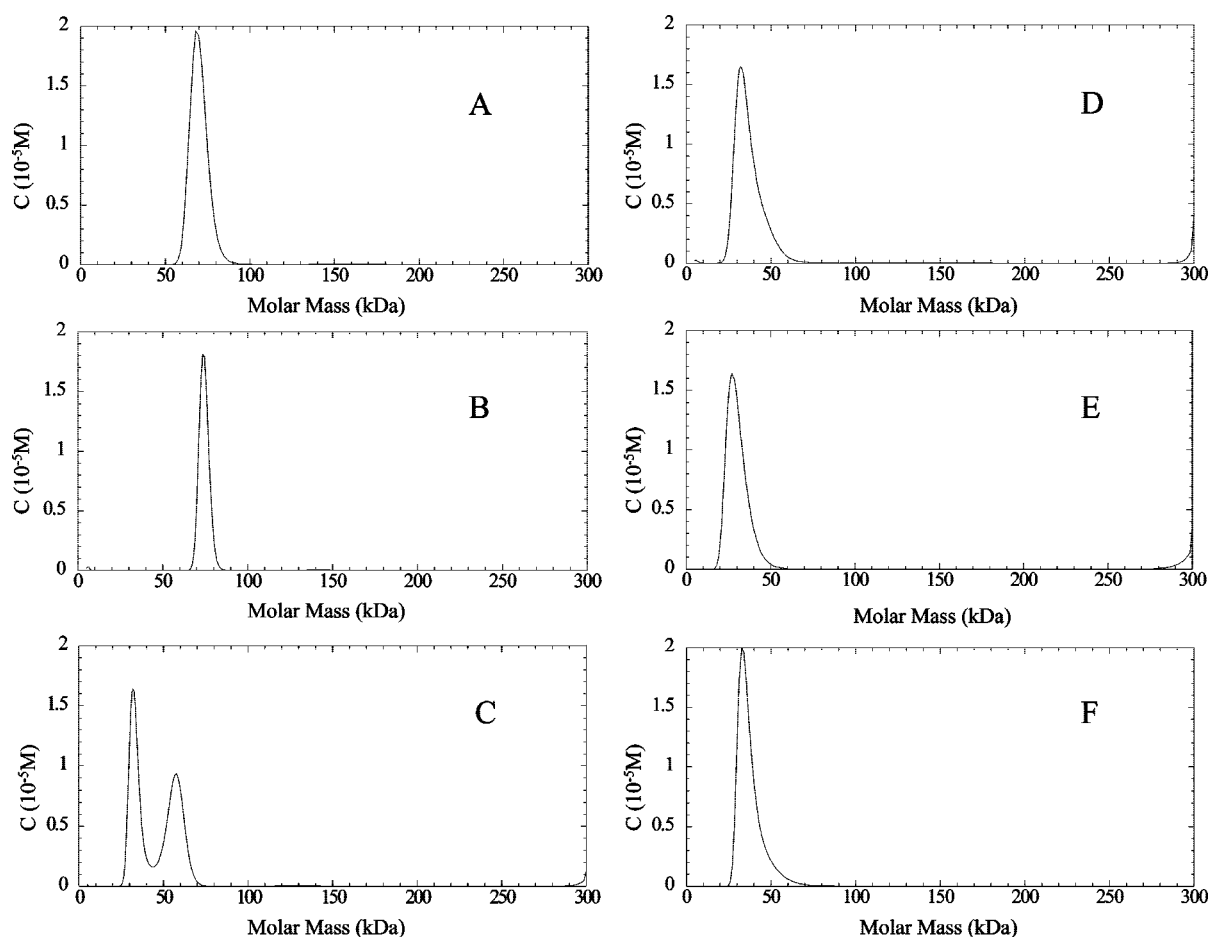
Figure S1 in the Supporting Information). Within each cluster, the number of coordinate sets was counted, and the coordinate set whose  $\Delta E_{\text{gas}}^{\text{elec}}$  is closest to the mean  $\Delta E_{\text{gas}}^{\text{elec}}$  of the cluster was chosen as the representative dimer coordinate set. As the monomer structure for the wild-type or double mutant GGPPs is not available, the chain A coordinates from the 10 representative dimer coordinate sets were used for the monomer coordinate sets. Previous works (e.g., La Font et al.<sup>37</sup> and references therein) have also used the structure of the unbound protein from that of the respective complex. Since the different clusters contain different numbers of coordinate sets, the  $\Delta G_{\text{sln}}$  values derived from the 10 representative coordinate sets were weighted according to the number of coordinate sets in

each cluster to give a final weighted average  $\Delta G_{\text{sln}}$  and respective standard deviation. For each of the 10 clusters, the per-residue contribution to the dimerization free energy,  $\Delta G_{\text{sln}}(i)$ , was also computed based on the representative coordinate set and assumed to be the same for all coordinate sets belonging to the same cluster. The standard deviation of the 1000  $\Delta G_{\text{sln}}(i)$  values yields the uncertainty/error in the calculated  $\Delta G_{\text{sln}}(i)$ .

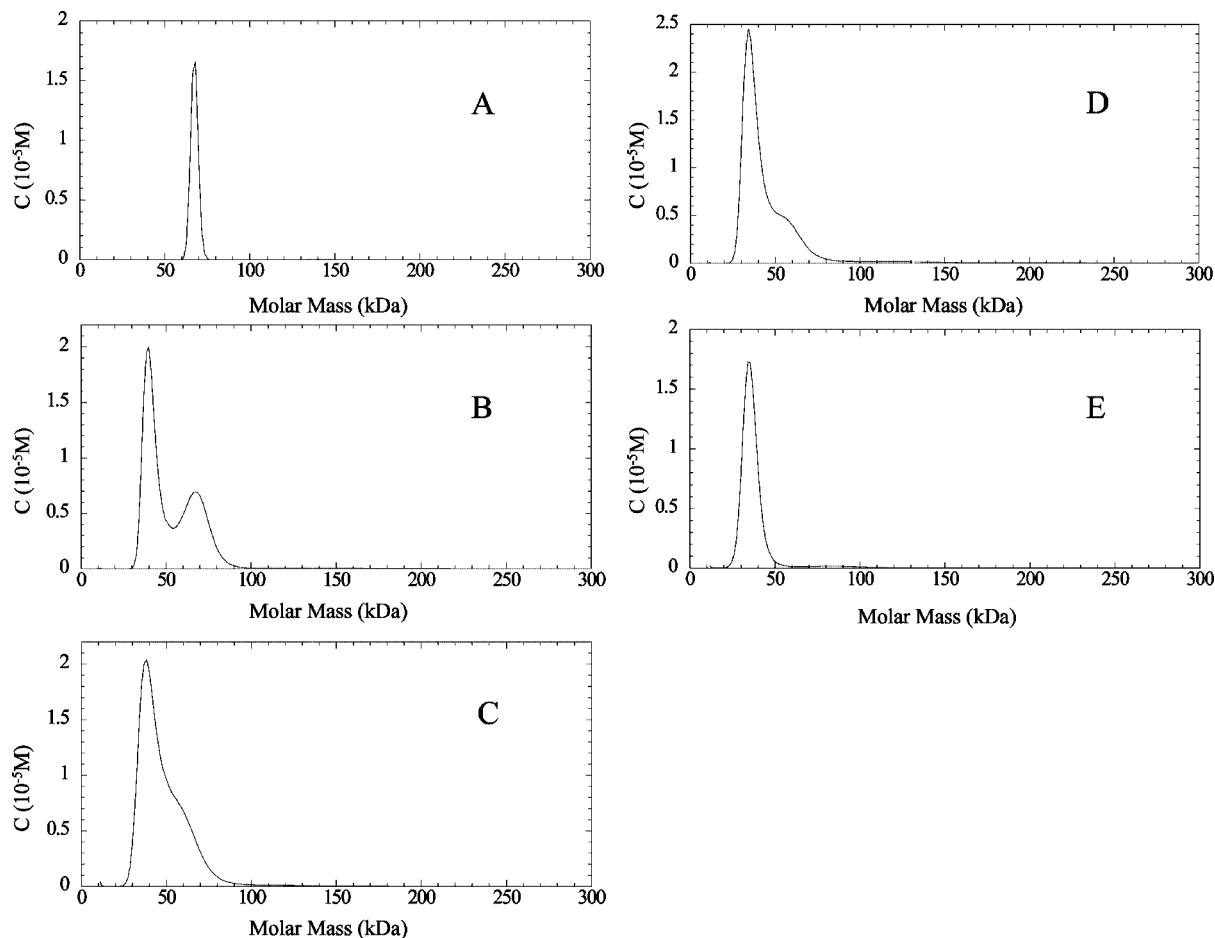
The motivation for using the above procedure, rather than calculating the free energies for all 1000 coordinate sets is (i) to speed up the free energy decomposition calculations and (ii) to ensure that the most frequently occurring configurations contribute more to the final free energy than the rarer configurations, which are also represented, albeit with less weight. All simulations and free energy decomposition calculations were carried out on a parallel cluster of Intel Core2 dual core CPUs with the GNU/Linux operating system.

## Results

**Expression and Purification of Mutant GGPPs.** Wild-type and the mutant GGPPs including those lacking the first couple of aa residues  $\Delta(1-6)$ ,  $\Delta(1-7)$ ,  $\Delta(1-8)$ ,  $\Delta(1-9)$ , and  $\Delta(1-17)$  as well as the site-directed mutants E7G, L8G, I9G, E7G/L8G, L8G/I9G, L163G, M167G, L200G, I203G, L163G/M167G, L200G/I203G, N101G, D145K, E134A, R175A, and M167G/N199A were expressed in *E. coli* and purified as previously described for the wild-type.<sup>21</sup> All the mutant enzymes were found in the soluble fraction of the cell lysate except L163G/M167G and L200G/I203G, which formed inclusion bodies. The



**Figure 2.** Quaternary structures of wild-type and the truncated mutants determined from AUC experiments. AUC data of (A) wild-type, (B)  $\Delta(1-6)$ , (C)  $\Delta(1-7)$ , (D)  $\Delta(1-8)$ , (E)  $\Delta(1-9)$ , and (F)  $\Delta(1-17)$  GGPPs are shown. The molecular mass of each mutant protein was deduced from the sedimentation velocity data. From the data, wild-type is a dimer,  $\Delta(1-6)$  is a mixture of dimer and monomer, whereas  $\Delta(1-8)$ ,  $\Delta(1-9)$ , and  $\Delta(1-17)$  are monomer.



**Figure 3.** Quaternary structures of single and double mutants determined from AUC data of (A) E7G, (B) L8G, (C) I9G, (D) E7G/L8G, and (E) L8G/I9G GGPPs. The molecular mass of each protein was deduced from the sedimentation velocity data. From the data, E7G/L8G is a mixture of dimer and monomer, I9G and E7G/L8G are mostly dimer with a small fraction of monomer, and L8G/I9G is a monomer.

soluble GGPPs had the yield of approximately 5–15 mg per liter of the cell culture medium after purification.

**L8 and I9 in Helix A are Essential for Dimerization.** To analyze the quaternary structures of the truncated and site-directed mutant GGPPs, the raw AUC data were analyzed by continuous size distribution, which implemented a highly reliable model, as indicated by the homogeneous bitmap picture (not shown). All of the data were derived from an excellent matching curve of the original raw sedimentation data and the randomly distributed residual values. The results, summarized in Table 1, show that the wild-type GGPPs is dimeric with a sedimentation coefficient of 4.8 S and a molecular weight of 78 kDa (Figure 2A). When the first six residues were deleted from the N-terminus, the  $\Delta(1-6)$  mutant protein remained as a dimer (Figure 2B). However, when the first seven residues were deleted (Figure 2C), a proportion of monomer appeared. With further successive deletion of the N-terminal residues, the  $\Delta(1-8)$ ,  $\Delta(1-9)$ , and  $\Delta(1-17)$  mutants showed a predominantly monomeric form (Figure 2D, E, and F, respectively). The data in Table 1 suggest that the residues essential for dimerization start from the seventh residue, E7.

To determine which of the aa residues are essential for dimerization, the site-directed mutants, E7G, L8G, I9G, E7G/

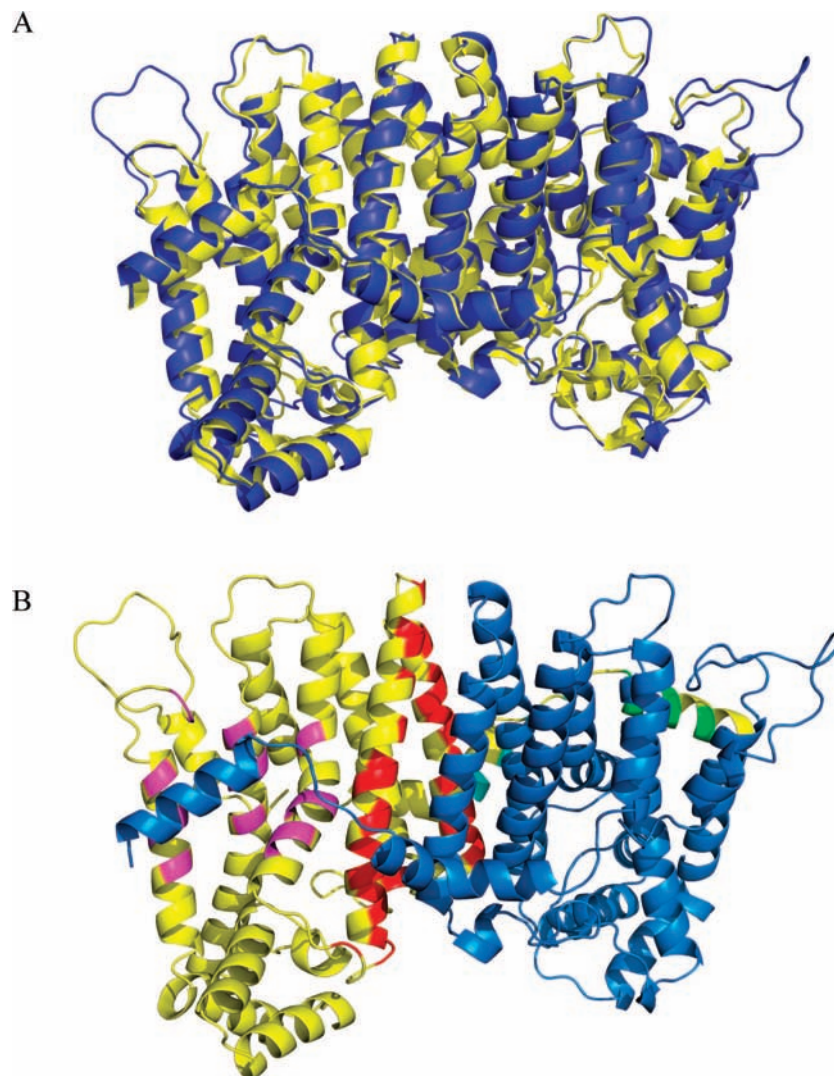
L8G, and L8G/I9G were constructed and analyzed using AUC. The E7G mutant remained as a dimer (Figure 3A), and the L8G mutant was a mixture of dimer and monomer (Figure 3B), whereas the I9G mutant was mainly a monomer (Figure 3C). The simultaneous mutations of E7 and L8 to Gly yielded nearly all monomers (Figure 3D), whereas the L8G/I9G double mutant was exclusively monomeric (Figure 3E). From the AUC data, the monomeric peaks for I9G (Figure 3C) and E7G/L8G (Figure 3D) were not completely symmetric, so a small fraction of dimer might exist; in contrast, the monomeric peak for L8G/I9G was symmetric, indicating a complete monomer with a sedimentation coefficient of 3.4 S and a molecular weight of 38 kDa. Therefore, L8 and I9 together provide critical interactions for maintaining the native dimeric state of GGPPs.

**Free Energy Changes upon Double Mutations of L8 and I9 to Glycine.** To verify that mutations of two nonpolar residues, L8 and I9, distal from the main dimer interface could destabilize the dimer, MD simulations of the wild-type GGPPs and the L8G/I9G double mutant dimers were performed, and the dimerization free energy difference between the wild-type and double mutant protein was computed, as described in the Materials and Methods section. Notably, simulation of the wild-type GGPPs dimer preserved the overall dimer structure as well as the key interactions across the protein–protein interface seen in the crystal structure. The rmsd of backbone helix atoms in the mean MD structure (see Materials and Methods) of the wild-type GGPPs dimer from those in the respective X-ray structure

(36) Hendsch, Z. S.; Tidor, B. *Protein Sci.* **1999**, *8*, 1381–1392.

(37) Lafont, V.; Schaefer, M.; Stote, R. H.; Altschuh, D.; Dejaegere, A. *Proteins: Struct., Funct., Bioinfo.* **2007**, *67*, 418–434.

(38) DeLano, W. L.; DeLano Scientific LLC: Palo Alto, CA, 2008.



**Figure 4.** (A) Overlay of the crystal structure (in yellow) of the wild-type GGPPs dimer (PDB entry 2DH4) with the respective average MD structure (in blue) derived from the last 100 ps of the simulation after a translational-rotational fit over the backbone heavy atoms. The MD structure has a modeled loop between residues 310 and 322 that is not seen in the crystal structure. (B) The interface residues (see text for definition) in the wild-type simulation structure: chain A is depicted in yellow, and chain B is displayed in light blue. For the chain A interface residues, those in helix A are in green, those linking helix A and the rest of the chain A are in yellow, and those comprising the major interface are in red, while those comprising the minor interface are in magenta.

after a translational-rotational fit over the backbone heavy atoms is 1.5 Å (Figure 4A). Most of the close contacts (atom–atom interactions  $\leq 4$  Å) across the dimer interface seen in the X-ray structure are maintained in the simulation (Table S1 of the Supporting Information).

As the dimerization free energy difference between the wild-type and double mutant GGPPs neglected conformational entropy contributions and was estimated using gas-phase energies combined with an implicit solvent model, it should be considered qualitative, and not quantitative, especially since the respective experimental value is not available for comparison. However, the trends in the relative free energies are expected to be reliable, as systematic errors in the free energies of two similar systems differing only in 2 residues (L8 and I9 mutated to glycine) out of 340 are expected to cancel. Indeed, the computed dimerization free energy for the double mutant GGPPs is significantly higher than that for the wild-type protein (by  $\sim 22$  kcal/mol, Table 2). This implies that, relative to the wild-type dimer, the L8G/I9G double mutant dimer is quite unstable,

consistent with the experimental finding of an exclusively monomeric L8G/I9G double mutant.

To determine how two nonpolar residues, L8 and I9, distal from the main dimer interface disrupts dimerization, the dimerization free energy of the double mutant GGPPs relative to that of the wild-type protein,  $\Delta\Delta G_{\text{sim}}$ , was decomposed into contributions made by the interface residues, defined as any atom in chain A within 4 Å from any atom in chain B for more than 50% of the sampled 1000 configurations. These 58 interface residues are listed in Table S2 of the Supporting Information together with their relative dimerization free energy contributions. 54 of the 58 interface residues fall into the following four regions (see Figure 4B): (i) the helix A interface (in green), composed of 7 residues; (ii) the linker interface (in yellow), composed of 6 residues, connecting helix A to the rest of chain A; (iii) the major interface (in red), composed of 28 residues in helices F and G; (iv) the minor interface (in magenta), composed of 13 residues close to a helix A residue from chain B. Of the four remaining interface residues, E78 and I155, which are

**Table 2.** Dimerization Free Energy for the Double Mutant GGPPs Relative to the Wild-Type Dimer

contribution to $\Delta\Delta G_{\text{sln}}$ from residue $i$ (no. of residues) <sup>a</sup>	$\Delta\Delta G_{\text{sln}}(i)$ (kcal/mol)
all residues except Thr -3 (338)	22.4 ± 0.2
all interface residues (58) <sup>b</sup>	24.5 ± 0.7
helix A interface residues (7) <sup>c</sup>	2.1 ± 0.8
linker interface residues (6) <sup>d</sup>	1.3 ± 0.5
major interface residues (30) <sup>e</sup>	7.7 ± 0.8
minor interface residues (15) <sup>f</sup>	13.4 ± 0.4

<sup>a</sup> Contribution of residues in chain A to  $\Delta\Delta G_{\text{sln}}$ ; the number in parentheses is the number of residues contributing to  $\Delta\Delta G_{\text{sln}}$ . <sup>b</sup> Contribution from all interface residues in the wild-type chain A, defined as any atom in chain A within 4 Å from any atom in chain B for more than 50% of the sampled time in the wild-type GGPPs simulation. <sup>c</sup> Contribution from interface residues in helix A (1, 2, 5, 6, 8–10). <sup>d</sup> Contribution from interface residues (13, 15, 19, 22, 23, 25) linking helix A to the rest of chain A. <sup>e</sup> Contribution from interface residues in helix D (78), helix F (97–98, 100–101, 104–105, 108–109, 111–112, 115), helix G (121, 124–125, 128, 132, 135–137, 139–140, 142–144, 146–147, 150–151), and connecting loop, I155. <sup>f</sup> Contribution from interface residues near helix A atoms from chain B in helix H (163–164, 167–168), helix I (175), helix J (196, 199–200, 203), helix O (291, 294–295, 298), 306, and 307.

**Table 3.** Residues with Large Changes in Dimerization Affinity per Heavy Atom, DA, upon Double Mutations of L8 and I9 to Glycine

residue $i$	interface	$\Delta\text{DA}(i)^a$ (kcal/mol)
I9	helix A	0.60
L8	helix A	0.58
Y307	minor	0.38
M167	minor	0.36
E121	major	0.35
E160	helix H	0.32
K25	linker	0.30
E7	helix A	0.28
N132	major	0.26
M298	minor	0.25

<sup>a</sup> Difference between the dimerization free energy contribution of residue  $i$  divided by its number of non-hydrogen atoms in the wild-type protein relative to that in the L8G/I9G double mutant.

within 3.5 Å from helix G atoms, were assigned to the major interface, whereas N168 and R175, which are within 3.5 Å from minor interface residues, M167 and N199, respectively, were assigned to the minor interface.

Interestingly, the double-site mutations of L8 and I9 to glycine appear to affect the major and minor interfaces more than helix A or its adjacent linker interface. The free energy change for the minor ( $\Delta\Delta G_{\text{sln}}(i) = 13.4$  kcal/mol) and the major ( $\Delta\Delta G_{\text{sln}}(i) = 7.7$  kcal/mol) interface residues are significantly more unfavorable than that for helix A ( $\Delta\Delta G_{\text{sln}}(i) = 2.1$  kcal/mol) or the linker ( $\Delta\Delta G_{\text{sln}}(i) = 1.3$  kcal/mol) interface residues (see Table 2). Thus, the L8G/I9G double mutations probably result in a loss of native major and minor interface contacts, inhibiting dimerization (see below).

Since the helix A interface residues make only a small net contribution to the impaired dimerization ability of the L8G/I9G double mutant, do the mutant residues contribute at all to the observed loss in dimerization? To address this question, we computed the dimerization affinity per heavy atom, DA, defined as the dimerization free energy contribution of each residue divided by its number of nonhydrogen atoms.<sup>37</sup> The DA values for the wild-type and double mutant proteins are listed in Table S2 of the Supporting Information, and those residues with DA changes  $\geq 0.25$  kcal/mol are listed in Table 3. Interestingly, among all the residues in the wild-type protein, I9 and L8 exhibit the most unfavorable DA change ( $\sim 0.60$  kcal/mol), and E7 is

**Table 4.** Changes in the Free Energy Dimerization Components for L8 and I9 upon Double Mutations to Glycine

residue $i$	$\Delta\Delta G_{\text{sln}}(i)$ (kcal/mol)	$\Delta\Delta E_{\text{gas}}^{\text{vdW}}(i)$ (kcal/mol)	$\Delta\Delta E_{\text{gas}}^{\text{elec}}(i)$ (kcal/mol)	$\Delta\Delta G_{\text{soln}}^{\text{elec}}(i)$ (kcal/mol)	$\Delta\Delta G_{\text{soln}}^{\text{nonelec}}(i)$ (kcal/mol)
L8	2.3 ± 0.3	2.5 ± 0.3	-0.4 ± 0.4	0.7 ± 0.8	-0.5 ± 0.3
I9	4.3 ± 0.3	4.7 ± 0.3	2.3 ± 0.4	-2.0 ± 0.6	-0.7 ± 0.2

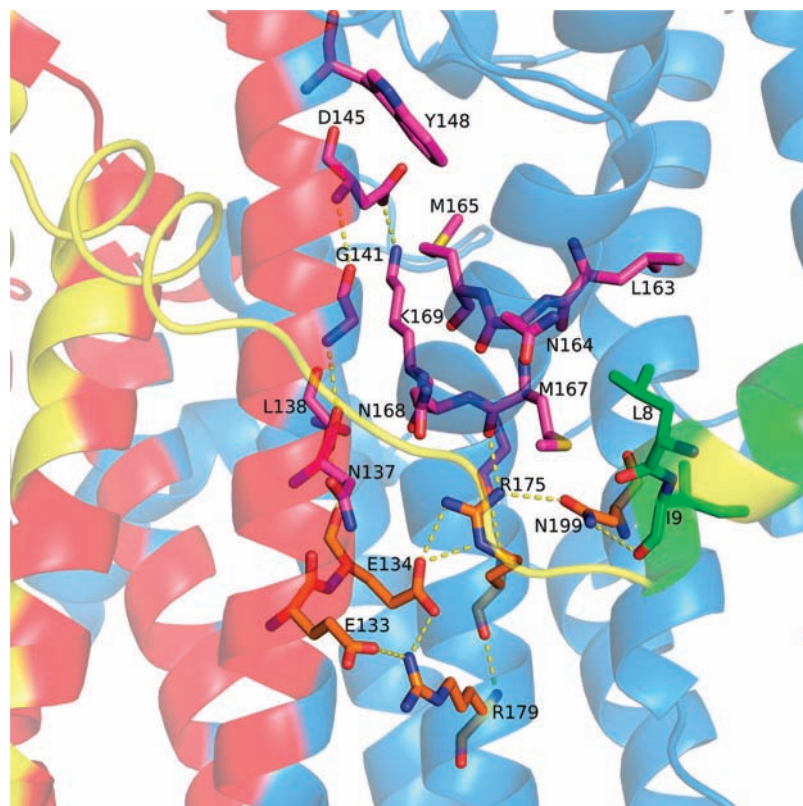
the only other residue in helix A that exhibits a significant unfavorable DA change (0.28 kcal/mol). These findings are in accord with the above experimental findings that residues essential for dimerization start from the seventh residue, and I9 and L8 are critical for dimerization. The other residues with unfavorable DA changes are located mostly at the major interface (E121, N132) and the minor interface (M167, M298, Y307).

**Interaction Changes Caused by the L8G/I9G Double Mutations.** To elucidate why mutations at positions 8 and 9 lead to unfavorable free energy contributions, the respective  $\Delta\Delta G_{\text{sln}}$  values were decomposed into the individual free energy components; that is,  $\Delta\Delta G_{\text{sln}}(i) = \Delta\Delta E_{\text{gas}}^{\text{vdW}}(i) + \Delta\Delta E_{\text{gas}}^{\text{elec}}(i) + \Delta\Delta G_{\text{soln}}^{\text{elec}}(i) + \Delta\Delta G_{\text{soln}}^{\text{nonelec}}(i)$  (see Materials and Methods). The results in Table 4 show that the impaired dimerization ability of the L8G/I9G double mutant is due mainly to the loss of native vdW contacts of L8 and I9 upon mutation to glycine, as the  $\Delta\Delta E_{\text{gas}}^{\text{vdW}}$  values (2.5 and 4.7 kcal/mol) are more positive than the other free energy components at these two positions. It is also due to the loss of electrostatic interactions of I9 upon mutation to glycine, as evidenced by the positive  $\Delta\Delta E_{\text{gas}}^{\text{elec}}$  (2.3 kcal/mol).

Comparison of the vdW and hydrogen-bonding interactions in the wild-type and the L8G/I9G dimer simulations shows that the double L8G/I9G mutations result in a direct loss of vdW and hydrogen-bonding interactions with residues at the minor interface. In the wild-type dimer simulation, L8 and I9 in chain A made extensive vdW contacts (heavy atom–heavy atom distance between 3.5 and 4 Å) with the following residues in chain B at the minor interface; namely, L163, N164, M167, P196, N199, L200, I203, and Y307 (see Figure 5). Furthermore, the I9 backbone O formed a hydrogen bond with the N199 side chain N<sup>δ2</sup> in the other subunit that was not found in the X-ray structure (see Table S1 of the Supporting Information). This may be due to a wrong assignment of the N199 side chain O<sup>δ1</sup> and N<sup>δ2</sup> in the X-ray structure, as their distances to I9 O are 2.99 and 5.19 Å, respectively. (In other words, O<sup>δ1</sup> in the X-ray structure should be N<sup>δ2</sup>). In contrast to the wild-type dimer simulation, the two Gly side chains had virtually no vdW contacts with the other subunit during the L8G/I9G simulation, resulting in positive  $\Delta\Delta E_{\text{gas}}^{\text{vdW}}$  values (Table 4). Moreover, unlike the wild-type GGPPs simulation structure, the mutant G9 did not form any hydrogen bonds in the simulation, so the loss of the wild-type hydrogen bond may account for the positive  $\Delta\Delta E_{\text{gas}}^{\text{elec}}$  (2.3 kcal/mol, Table 4). Thus, L8 and I9 directly contact minor interface residues, so their mutations to glycine result in a loss of vdW and hydrogen-bonding interactions, thus destabilizing the minor interface.

How do the L8G/I9G mutations also affect the major dimer interface, as evidenced by the unfavorable free energy contribution from the major interface residues (7.7 kcal/mol)? As shown in Figure 5, L8 makes vdW contacts with L163, N164, and M167, whose neighboring residues, M165, N168, K169, and G171, are within vdW contact of N137, L138, G141, and D145, located at/near the major dimer interface. In particular, K169 forms a salt-bridge with D145, which in turn forms a backbone-





**Figure 5.** Two plausible networks of vdW and/or hydrogen-bonding interactions linking L8 and I9 to the major dimer interface deduced from MD simulations. The residues with carbons depicted in magenta link L8 to the major dimer interface via vdW contacts with L163, N164, and M167, whose neighboring residues, M165, N168, K169, and G171, are within vdW contact of N137, L138, G141, D145, and Y148, located at/near the major dimer interface. The residues with carbons depicted in orange link I9 to the major dimer interface via a hydrogen-bonding network involving N199, R175, E134, R179, and E133. These residues are shown in stick with the oxygen and nitrogen atoms in red and blue, respectively. Chain A is depicted in yellow, chain B is displayed in light blue, chain A helix A interface residues are in green, and both chain A and chain B major interface residues are red.

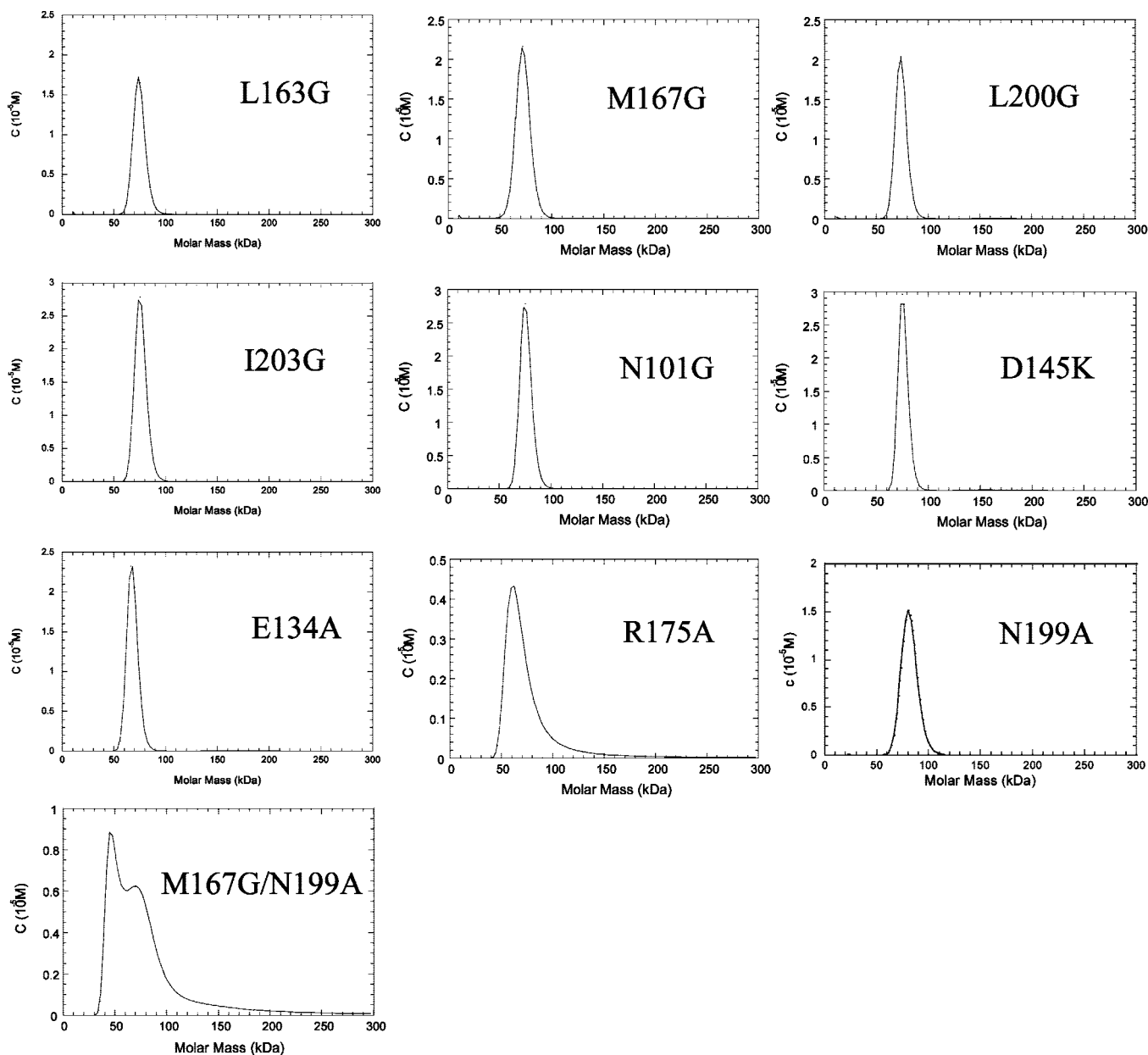
backbone hydrogen bond with G141, which itself is hydrogen bonded to N137, a major interface residue. Likewise, I9 is hydrogen bonded to N199, which in turn is hydrogen bonded to R175, which forms a salt-bridge with E134 located near the major dimer interface. Thus, via a network of vdW contacts and hydrogen bonds, L8 and I9 in helix A may also affect the more distal major interface.

**Interactions Mediating Dimerization Probed by Site-Directed Mutagenesis.** Since two networks of interactions, one stemming from L8 and the other from I9, stabilize dimerization, these interactions were probed by creating single mutants, L163G, M167G, L200G, I203G, N101G, D145K, E134A, R175A, and three double mutants, M167G/N199A, L163G/M167G, and L200G/I203G. The results in Figure 6 show that all the single mutants with only one residue downstream of the L8 and I9 networks altered remained as dimers. In contrast, the double mutants, L163G/M167G and L200G/I203G, formed inclusion bodies, indicating complete unfolding of the mutant protein. Thus, the packing interactions of L163, M167, L200, and I203 appear to play an important role in maintaining the protein structure and stability. On the other hand, the M167G/N199A double mutant became a mixture of monomer/dimer. This indicates that the disruption of the GGPPs dimer into monomer stems from the L8G/I9G mutations.

**GGPPs Dimeric Structure is Essential for Enzyme Function.** To examine the correlation between quaternary structures and enzyme function, the enzymatic activities of the truncated and site-directed mutants were measured and summarized in Table 5. The dimeric forms of the  $\Delta(1-6)$  and E7G

proteins retained 70–80% activity of the wild-type. As the  $\Delta(1-7)$  and L8G mutants have slightly more monomeric than dimeric proteins (dimer dissociation constant  $K_d = 1.5$  and  $2.0 \mu\text{M}$ , respectively), they showed a dramatic decrease in activity (0.3- and 0.2-fold, respectively). Monomeric mutants including  $\Delta(1-8)$ ,  $\Delta(1-9)$ , and L8G/I9G displayed only residual activities, approximately  $10^{-3}$ -fold of the wild-type activity. However, the monomeric  $\Delta(1-17)$  showed no detectable enzyme activity. Notably, the monomeric and partially monomeric mutants only showed decreased  $k_{\text{cat}}$  values, but the  $K_m$  values remained unchanged. It could not be ruled out that a very small fraction of dimer was present in the “monomeric” mutants, which was undetectable by AUC. The results in Table 5 show that the GGPPs dimeric structure is essential for optimal catalytic activity.

**Possible Causes of Activity Loss in Monomeric GGPPs.** From the crystal structure,<sup>20</sup> the GGPPs contains an active site in each monomer, but the active site is formed not only by residues in this monomer but also by residues from the other monomer. As shown in Figure 7A top panel, three residues from chain B (shown in yellow) are involved in forming the active site in chain A (shown in blue). However, without the three B-chain residues, a puncture is formed in the center (Figure 7A bottom panel). The B-chain residues, N101, N104, and Y105 (shown in yellow), form hydrogen bonds with the A-chain residues, H139, N104, and R140 (shown in blue), respectively, in the active site of chain A (see Figure 7B). In chain A, H139 and R140 are linked via a hydrogen bond with G143, which in turn is hydrogen bonded to Q142, which is directly and indirectly



**Figure 6.** AUC analysis of mutant GGPPs. From AUC experiments, L163G, M167G, L200G, I203G, N101G, D145K, E134A, and R175A are dimer. M167G/N199A double mutant is a mixture of dimer and monomer.

(via E78) hydrogen bonded to D75. D75 is directly hydrogen bonded with the substrate FPP and indirectly hydrogen bonded with the other substrate IPP via hydrogen-bonding interactions with L72 and H68. Hence, without the three residues from chain B, the interaction network for catalysis may be disrupted, so the monomeric mutant showed none or dramatically reduced activity.

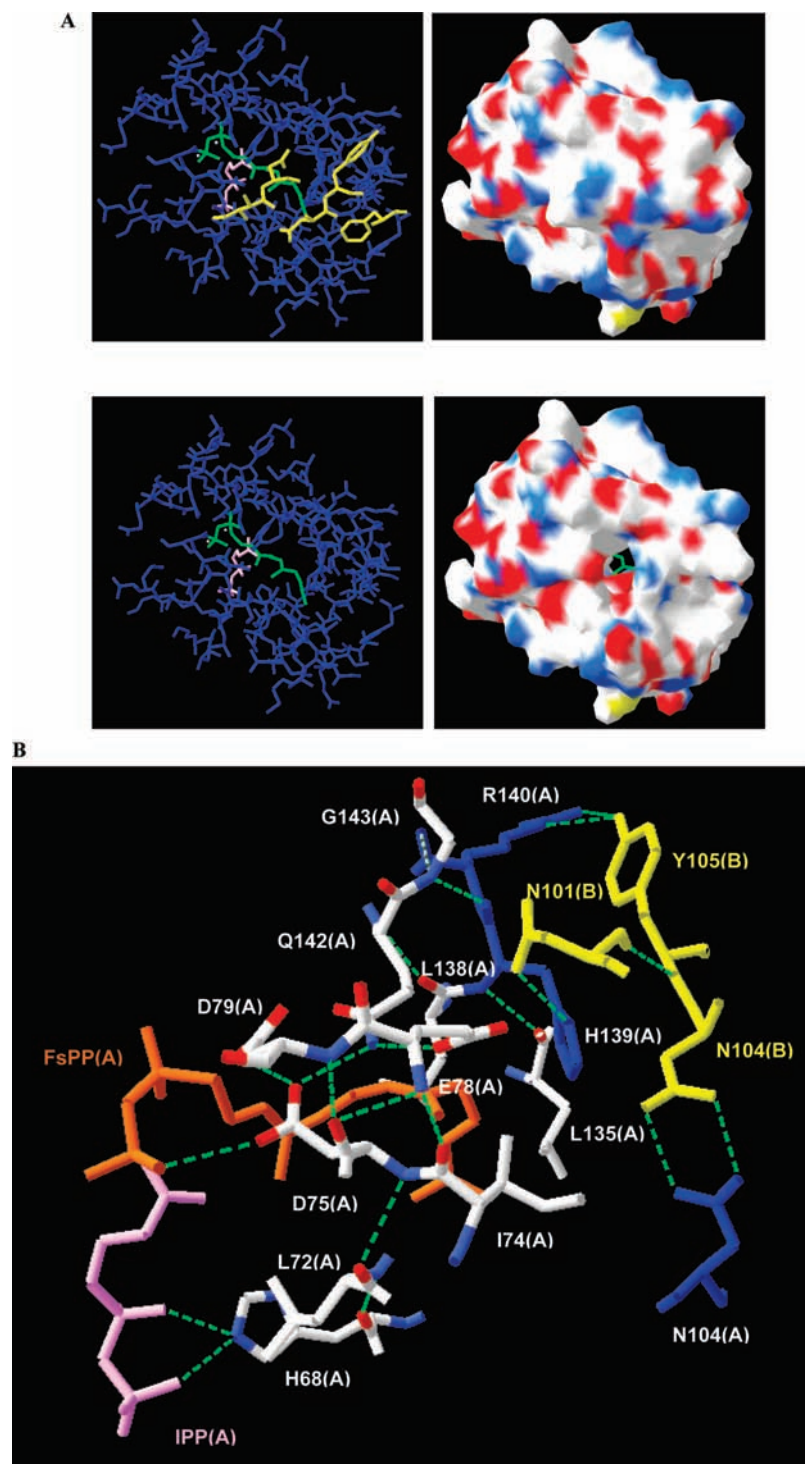
## Discussion

There is increasing evidence that dimerization is required for enzyme activity, especially for proteases. The reason is obvious for HIV protease, since two monomers together form an active site by providing a catalytic Asp residue from each monomer. Here, we present an interesting case where mutations of only two residues, L8 and I9, in the first N-terminal helix, far away from the main dimer interface, can completely disrupt a dimeric *S. cerevisiae* enzyme GGPPs into a monomer with at least  $10^3$ -fold lower activity.

**Table 5.** Kinetic Parameters and Dimer Dissociation Constant ( $K_d$ ) of Wild-Type and Mutant GGPPs

yeast GGPPs	$K_m^{FPP}$ ( $\mu\text{M}$ )	$K_m^{IPP}$ ( $\mu\text{M}$ )	$k_{cat}$ ( $\text{s}^{-1}$ )	activity fold	$K_d^a$ ( $\mu\text{M}$ )
wild-type	$3.2 \pm 0.3$	$0.8 \pm 0.2$	$(2.5 \pm 0.4) \times 10^{-2}$	1	N.A. (D)
$\Delta(1-6)$	$4.3 \pm 0.6$	$0.6 \pm 0.1$	$(1.8 \pm 0.2) \times 10^{-2}$	0.72	N.A. (D)
$\Delta(1-7)$	$1.4 \pm 0.3$	$0.4 \pm 0.3$	$(7.3 \pm 0.4) \times 10^{-3}$	0.29	1.47 (M + D)
$\Delta(1-8)$	$2.6 \pm 0.7$	$0.5 \pm 0.2$	$(6.6 \pm 0.7) \times 10^{-5}$	0.003	N.A. (M)
$\Delta(1-9)^b$	$1.9 \pm 0.3$	$1.4 \pm 0.5$	$(4.2 \pm 0.3) \times 10^{-5}$	0.002	N.A. (M)
$\Delta(1-17)^b$	N.A.	N.A.	N.A.	N.A.	N.A. (M)
E7G	$4.4 \pm 0.8$	$0.4 \pm 0.2$	$(2.1 \pm 0.1) \times 10^{-2}$	0.84	N.A. (D)
L8G	$2.4 \pm 0.1$	$0.5 \pm 0.1$	$(5.1 \pm 0.2) \times 10^{-3}$	0.2	2.04 (M + D)
I9G	$3.6 \pm 0.2$	$0.7 \pm 0.3$	$(6.2 \pm 0.3) \times 10^{-4}$	0.025	N.A. <sup>c</sup>
E7G/L8G	$2.8 \pm 0.3$	$0.3 \pm 0.2$	$(1.5 \pm 0.2) \times 10^{-4}$	0.006	N.A. <sup>c</sup>
L8G/I9G	$3.1 \pm 0.2$	$0.6 \pm 0.1$	$(7.2 \pm 0.4) \times 10^{-5}$	0.003	N.A. (M)

<sup>a</sup> For pure dimer or monomer,  $K_d$  can not be determined (N.A., not available). In parentheses, D means dimer and M means monomer. <sup>b</sup> Data from Chang et al.<sup>20</sup>  $\Delta(1-17)$  has no detectable activity. <sup>c</sup> For I9G and E7G/L8G, dimer and monomer were not well resolved.



**Figure 7.** (A) Active-site structures of GGPPs in chain A with and without chain B. In the top panel, FPP and IPP in the active site of chain A (A-chain residues are shown in blue) surrounded by the residues from chain B (shown in yellow). In the bottom panel, the active site in chain A without the residues from chain B shows a puncture in the center. (B) The detailed interactions to connect the three active-site residues N104(B), Y105(B), and N101(B) from chain B and the bound FPP and IPP in chain A. This figure is based on the crystal structure reported in ref 10.

Although helix A is distant from the main dimer interface (Figure 1), it contributes to the GGPPs subunit dimerization, as it stabilizes the minor interface via direct hydrogen bonds and vdW interactions. Helix A binds into a pocket formed by the other subunit residues, L163, N164, M167, N168, P196, N199, L200, I203, F291, Q294, L295, M298, K306, and Y307, which form the minor interface (see Figure 5 and Table S2 of the Supporting Information). The wild-type dimer simulation

shows that the helix A residues, D6 and I9, form hydrogen bonds with the K306/Y307 and N199 sidechains, respectively (see Table S1 of the Supporting Information). In addition to these hydrogen bonds, the helix A residues also make extensive vdW contacts with the pocket; namely, M1 with F291 and Q294, E2 with M298, I5 with L200 and M298, L8 with L163, N164, M167, and I203, and I9 with P196, N199, L200, I203, and Y307. Compared to the combined role of L8 and I9, the other

helix A residues, M1, E2, and I5, appear to play a secondary role in stabilizing the minor interface, as they make fewer vdW contacts with chain B residues between F291 and M298.

Helix A is also linked to the major interface via a contiguous network of vdW and hydrogen-bonding interactions. In particular, I9 is linked to the major interface via the following hydrogen-bonding network; namely, I9...N199...R175...R179...E133 (see Figure 5). I5 is indirectly linked to this hydrogen-bonding network, as it is in vdW contact with L200, the neighbor of N199. Like I9, L8 is linked to the major interface via a network of vdW and hydrogen-bonding interactions: L8 makes vdW contacts with M167, whose neighbor, K169, forms a salt-bridge with D145, which in turn forms backbone-backbone hydrogen bonds with G141 and N137 at the major interface (see Figure 5). The other helix A residues, M1 and E2, make vdW contacts with F291, Q294, and M298, located on the external protein surface and are not linked to the major interface.

Interestingly, mutations of only two nonpolar residues, L8 and I9, in helix A to glycine suffice to disrupt dimerization. The L8G/I9G simulation shows that the smaller Gly side chains cannot form interactions in the same way as the wild-type residues can with the pocket described above. In contrast to L8 and I9, the mutant G8 and G9 made virtually no vdW contacts or hydrogen bonds with the minor interface residues, thus making a significantly unfavorable dimerization free energy contribution compared to their contribution in the wild-type protein (see Table 2). Furthermore, the mutant G8 and G9 may lack the ability of their wild-type counterparts to stabilize the major interface via the network described above, resulting in an unfavorable free energy change for the major interface residues (see Table 2). Thus, the double mutations of L8 and I9 to glycine result in a loss of protein-protein interactions

across the minor interface, which in turn may affect the major interface, thus destabilizing the dimeric state.

In summary, we have determined two residues (L8 and I9) that are essential in stabilizing the entire dimer of yeast GGPPs and have rationalized their roles in forming a network of vdW and hydrogen-bonding interactions with both minor and major interface residues using MD simulations. Since M167G remained a dimer, whereas M167G/N199A became a mixture of monomer/dimer (Figure 6), the hydrogen bond between I9 and N199 may play a more important role in maintaining the dimerization than vdW contacts between L8 and M167. In fact, the single mutant I9G is nearly monomeric, with only a small dimeric fraction. However, the mutation of L8 to glycine is required in addition to the mutation of I9 to glycine to yield a complete monomer. The resulting L8G/I9G monomer lacking the three residues from the other monomer in the active site (Figure 7) displays abolished activity. Our studies provide a better understanding of dimerization and activity modulation by long-range interactions.

**Acknowledgment.** We thank Dr. Hui-Chuang Chang for technical assistance in the AUC experiments and Tao-Hsin Chang for preparing Figure 1. This work was supported by the National Science Council, Taiwan (NSC contract no. 97-14-B-MY5 to C.L. and NSC contract no. 96-2113-M-001-016-MY3 to P.H.L.).

**Supporting Information Available:** Figure illustrating the clustering of the 1000 coordinate sets from the wild-type GGPPs simulation, table showing comparison of close contacts in X-ray structure of GGPPs, and table of  $\Delta G_i$  and DA values for interface residues in wild-type and double mutant GGPPs. This material is available free of charge via the Internet at <http://pubs.acs.org>.

JA808699C