

3-28-2010

## **Globular Structure of a Human Immunodeficiency Virus-1 Protease (1DIFA dimer) in an Effective Solvent Medium by a Monte Carlo Simulation**

Ras B. Pandey

*University of Southern Mississippi, ras.pandey@usm.edu*

Barry L. Farmer

*Wright Patterson Air Force Base, barry.farmer@wpafb.af.mil*

Follow this and additional works at: [https://aquila.usm.edu/fac\\_pubs](https://aquila.usm.edu/fac_pubs)



Part of the [Astrophysics and Astronomy Commons](#)

---

### **Recommended Citation**

Pandey, R. B., Farmer, B. L. (2010). Globular Structure of a Human Immunodeficiency Virus-1 Protease (1DIFA dimer) in an Effective Solvent Medium by a Monte Carlo Simulation. *Journal of Chemical Physics*, 132(12).

Available at: [https://aquila.usm.edu/fac\\_pubs/767](https://aquila.usm.edu/fac_pubs/767)

This Article is brought to you for free and open access by The Aquila Digital Community. It has been accepted for inclusion in Faculty Publications by an authorized administrator of The Aquila Digital Community. For more information, please contact [aquilastaff@usm.edu](mailto:aquilastaff@usm.edu).

# Globular structure of a human immunodeficiency virus-1 protease (1DIFA dimer) in an effective solvent medium by a Monte Carlo simulation

R. B. Pandey<sup>1,a)</sup> and B. L. Farmer<sup>2</sup>

<sup>1</sup>*Department of Physics and Astronomy, University of Southern Mississippi, Hattiesburg, Mississippi 39406-5046, USA*

<sup>2</sup>*Materials and Manufacturing Directorate, Air Force Research Laboratory, Wright Patterson Air Force Base, Ohio 45433, USA*

(Received 15 January 2010; accepted 17 February 2010; published online 22 March 2010)

A coarse-grained model is used to study the structure and dynamics of a human immunodeficiency virus-1 protease (1DIFA dimer) consisting of 198 residues in an effective solvent medium on a cubic lattice by Monte Carlo simulations for a range of interaction strengths. Energy and mobility profiles of residues are found to depend on the interaction strength and exhibit remarkable segmental symmetries in two monomers. Lowest energy residues such as Arg<sup>41</sup> and Arg<sup>140</sup> (most electrostatic and polar) are not the least mobile; despite the higher energy, the hydrophobic residues (Ile, Leu, and Val) are least mobile and form the core by pinning down the local segments for the globular structure. Variations in the gyration radius ( $R_g$ ) and energy ( $E_c$ ) of the protein show nonmonotonic dependence on the interaction strength with the smallest  $R_g$  around the largest value of  $E_c$ . Pinning of the conformations by the hydrophobic residues at high interaction strength seems to provide seed for the protein chain to collapse. © 2010 American Institute of Physics. [doi:10.1063/1.3358340]

The catalytic function of an enzyme such as 1DIFA, a human immunodeficiency virus (HIV)-1 protease, via hydrolysis of peptide bonds<sup>1</sup> is believed to be critical for the growth of HIV in host cells. Despite a considerable interest in recent years,<sup>2–8</sup> the search for an ideal inhibitor is far from complete due to mutation of HIV protease. Understanding the structure and dynamics of such a protein in an appropriate solvent is therefore very important in the search for protease inhibitors—a targeted effort toward finding ways to curtail the progression of HIV infection. Structural stability of HIV-1 protease has been extensively examined using all atom<sup>6</sup> molecular dynamics (MD) and coarse-grained<sup>9</sup> MD simulations in an effort to understand the effect of inhibitors. Most of the analysis seems to concentrate around an astoundingly symmetric structure from the protein data bank (PDB) involving beta sheets,  $\alpha$  helices, and strands resembling the elements of a face involving flaps with ears and eyes.<sup>1,6,9</sup> Recovering such a structure in a coarse-grained model is amazing at such a microscopic detail, i.e., a specific conformation from a huge ensemble with a large number of configurations in computer simulation models where invoking approximations is inevitable. Of course, a x-ray snapshot of the structure does not tell much about all the possible conformations, nor do snapshots from computer simulations guarantee the results from approximate force field, constraints, or *ad hoc* assumptions. Yet, comparison of the quantitative results of computer simulations<sup>9</sup> with the experimental snaps may be possible but somewhat premature or accidental. It is not feasible to explore the complete conformational phase space via the all-atom MD simulations. All-atom simulations involve some level of coarse graining in

incorporating the effects of electrons in atoms. Reproducing a structure in any simulation (coarse-grained or all-atom) is perhaps accidental or a result of constrained exploration around a limited phase space. We constrain our goal toward gaining a general insight and predicting qualitative trends in our coarse-grained model.

The last few decades have witnessed an enormous surge in interest<sup>10–29</sup> in modeling proteins. Much attention is focused on identifying the universal characteristics, e.g., folding pathways via analysis of the energy landscapes of protein chains as well as their specific characteristics that entail local structures to understand binding to pertinent targets. While all-atom modeling is generally appropriate for probing local structures in short time, coarse graining is almost unavoidable to address large-scale properties. Using a coarse-grained Monte Carlo (MC) simulation we have recently examined<sup>30</sup> both local and global properties such as energy and mobility profiles, structural correlations of each residue, and conformation of a HIV-1 protease (1DIFA monomer) which consists of 99 residues. We would like to extend this study to a 1DIFA dimer formed by combining the C-terminal residues (96–99) of one monomer with the N-terminal (residues 1–4) of the second and see if there are significant changes in properties.

TABLE I. A typical interaction matrix among  $H$ ,  $P$ ,  $E$ , and  $W$  components.

	$H$	$P$	$E$	$W$
$H$	$\epsilon_{HH}=0.0$	$\epsilon_{HP}=0.0$	$\epsilon_{HE}=0.0$	$\epsilon_{HW}=0.1$
$P$	$\epsilon_{PH}=0.0$	$\epsilon_{PP}=-0.2$	$\epsilon_{PE}=-0.2$	$\epsilon_{PW}=-0.2$
$E$	$\epsilon_{EH}=0.0$	$\epsilon_{EP}=-0.2$	$\epsilon_{EE}$	$\epsilon_{EW}=-0.3$
$W$	$\epsilon_{WH}=0.1$	$\epsilon_{WP}=-0.2$	$\epsilon_{WE}=-0.3$	$\epsilon_{WW}=-0.1$

<sup>a)</sup>Electronic mail: ras.pandey@usm.edu.

# Report Documentation Page

Form Approved  
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE <b>JAN 2010</b>		2. REPORT TYPE		3. DATES COVERED <b>00-00-2010 to 00-00-2010</b>	
4. TITLE AND SUBTITLE <b>Globular structure of a human immunodeficiency virus-1 protease (1DIFA dimer) in an effective solvent medium by a Monte Carlo simulation</b>				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) <b>Air Force Research Laboratory, Materials and Manufacturing Directorate, Wright Patterson AFB, OH, 45433</b>				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT <b>Approved for public release; distribution unlimited</b>					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
<b>unclassified</b>	<b>unclassified</b>	<b>unclassified</b>	<b>Same as Report (SAR)</b>	<b>5</b>	

## I. SEQUENCE OF 1DIFA DIMER (HIV-1 PROTEASE)

P<sup>1</sup>Q<sup>2</sup>T<sup>3</sup>R<sup>4</sup>L<sup>5</sup>W<sup>6</sup>Q<sup>7</sup>R<sup>8</sup>P<sup>9</sup>L<sup>10</sup>V<sup>11</sup>T<sup>12</sup>I<sup>13</sup>K<sup>14</sup>I<sup>15</sup>G<sup>16</sup>G<sup>17</sup>Q<sup>18</sup>L<sup>19</sup>K<sup>20</sup>E<sup>21</sup>A<sup>22</sup>L<sup>23</sup>L<sup>24</sup>D<sup>25</sup>T<sup>26</sup>G<sup>27</sup>A<sup>28</sup>D<sup>29</sup>D<sup>30</sup>T<sup>31</sup>V<sup>32</sup>L<sup>33</sup>E<sup>34</sup>E<sup>35</sup>M<sup>36</sup>S<sup>37</sup>  
 L<sup>38</sup>P<sup>39</sup>G<sup>40</sup>R<sup>41</sup>W<sup>42</sup>K<sup>43</sup>P<sup>44</sup>K<sup>45</sup>M<sup>46</sup>T<sup>47</sup>G<sup>48</sup>G<sup>49</sup>I<sup>50</sup>G<sup>51</sup>F<sup>52</sup>I<sup>53</sup>S<sup>54</sup>K<sup>55</sup>V<sup>56</sup>R<sup>57</sup>Q<sup>58</sup>Y<sup>59</sup>D<sup>60</sup>Q<sup>61</sup>I<sup>62</sup>L<sup>63</sup>I<sup>64</sup>E<sup>65</sup>I<sup>66</sup>C<sup>67</sup>G<sup>68</sup>H<sup>69</sup>K<sup>70</sup>A<sup>71</sup>I<sup>72</sup>  
 G<sup>73</sup>T<sup>74</sup>V<sup>75</sup>L<sup>76</sup>V<sup>77</sup>G<sup>78</sup>P<sup>79</sup>T<sup>80</sup>P<sup>81</sup>V<sup>82</sup>N<sup>83</sup>I<sup>84</sup>I<sup>85</sup>G<sup>86</sup>R<sup>87</sup>N<sup>88</sup>L<sup>89</sup>L<sup>90</sup>T<sup>91</sup>Q<sup>92</sup>I<sup>93</sup>G<sup>94</sup>C<sup>95</sup>T<sup>96</sup>L<sup>97</sup>N<sup>98</sup>F<sup>99</sup>P<sup>100</sup>Q<sup>101</sup>I<sup>102</sup>T<sup>103</sup>L<sup>104</sup>W<sup>105</sup>  
 Q<sup>106</sup>R<sup>107</sup>P<sup>108</sup>L<sup>109</sup>V<sup>110</sup>T<sup>111</sup>I<sup>112</sup>K<sup>113</sup>I<sup>114</sup>G<sup>115</sup>G<sup>116</sup>Q<sup>117</sup>L<sup>118</sup>K<sup>119</sup>E<sup>120</sup>A<sup>121</sup>L<sup>122</sup>L<sup>123</sup>D<sup>124</sup>T<sup>125</sup>G<sup>126</sup>A<sup>127</sup>D<sup>128</sup>D<sup>129</sup>T<sup>130</sup>V<sup>131</sup>L<sup>132</sup>  
 E<sup>133</sup>E<sup>134</sup>M<sup>135</sup>S<sup>136</sup>L<sup>137</sup>P<sup>138</sup>G<sup>139</sup>R<sup>140</sup>W<sup>141</sup>K<sup>142</sup>P<sup>143</sup>K<sup>144</sup>M<sup>145</sup>I<sup>146</sup>G<sup>147</sup>L<sup>148</sup>I<sup>149</sup>V<sup>150</sup>G<sup>151</sup>F<sup>152</sup>I<sup>153</sup>K<sup>154</sup>V<sup>155</sup>R<sup>156</sup>Q<sup>157</sup>Y<sup>158</sup>  
 D<sup>159</sup>Q<sup>160</sup>I<sup>161</sup>L<sup>162</sup>I<sup>163</sup>E<sup>164</sup>I<sup>165</sup>C<sup>166</sup>G<sup>167</sup>H<sup>168</sup>K<sup>169</sup>A<sup>170</sup>I<sup>171</sup>G<sup>172</sup>T<sup>173</sup>V<sup>174</sup>L<sup>175</sup>V<sup>176</sup>G<sup>177</sup>P<sup>178</sup>T<sup>179</sup>P<sup>180</sup>V<sup>181</sup>N<sup>182</sup>I<sup>183</sup>I<sup>184</sup>G<sup>185</sup>  
 R<sup>186</sup>N<sup>187</sup>L<sup>188</sup>L<sup>189</sup>T<sup>190</sup>Q<sup>191</sup>I<sup>192</sup>G<sup>193</sup>C<sup>194</sup>T<sup>195</sup>L<sup>196</sup>N<sup>197</sup>F<sup>198</sup>.

We consider the coarse-grained description<sup>30</sup> for the 1DIFA dimer, which consists of 198 residues tethered together on a cubic lattice. As described before,<sup>30</sup> the empty lattice sites act as solvent sites for effective medium. Each residue interacts via a Lennard-Jones (LJ) interaction

$$U_{ij} = f \varepsilon_{ij} \left[ \left( \frac{\sigma}{r_{ij}} \right)^{12} - \left( \frac{\sigma}{r_{ij}} \right)^6 \right], \quad r_{ij} < r_c$$

with surrounding residues and solvent (*W*) sites at distance  $r_{ij}$  within a range  $r_c = \sqrt{8}$ ;  $\sigma = 1$ . The interaction  $\varepsilon_{ij}$  is unique for each residue type. The parameter  $f$ , i.e., the depth of the LJ potential, can be varied to control the quality of the solvent. The characteristic of each residue is captured by incorporating its relative hydrophobic (*H*), polar (*P*), and electrostatic (*E*) strengths in normalizing their interaction<sup>30</sup> presented in Table I. Interactions ( $\varepsilon_{EE}$ ) among the electrostatic residues Asp ( $E_1$ ), Glu ( $E_2$ ), Lys ( $E_3$ ), and Arg ( $E_4$ ) are:  $\varepsilon_{11} = \varepsilon_{12} = \varepsilon_{22} = \varepsilon_{33} = \varepsilon_{34} = \varepsilon_{44} = 0.1$ ,  $\varepsilon_{13} = \varepsilon_{14} = \varepsilon_{23} = \varepsilon_{24} = -0.4$ . The magnitude of each interaction element ( $\varepsilon_{ij} = \varepsilon_{ji}$ ) is based on the insight gained from the all-atom description and known general characteristics.<sup>30</sup>

Attempts are made to move each residue randomly with the Metropolis algorithm subject to excluded volume constraints and the limits on changes in the covalent bond length.<sup>30</sup> For example, a randomly selected residue at a site  $i$  is moved to one of its randomly selected neighboring lattice sites, such as  $j$  with the Boltzmann probability  $\exp(-\Delta E_{ij}/T)$ , where  $\Delta E_{ij}$  is the change in energy between its new ( $E_j$ ) and old ( $E_i$ ) configuration  $\Delta E_{ij} = E_j - E_i$  and  $T$  is the temperature in units of the Boltzmann constant and the energy ( $f\varepsilon_{ij}$ ). The attempt to move each residue once defines one MC step time. A number of local and global physical quantities are evaluated during the simulations. Physical quantities in our study include energy of each residue, its mobility, average number of neighboring residues, and mean square displacement of the center of mass of the protein and its radius of gyration, etc. Simulations are performed for a sufficiently long time (typically for  $1 \times 10^7$  time steps) with many independent samples to estimate these quantities (see below). Different lattice sizes are used to test for the finite size effects on the qualitative variations in these physical quantities; most of the data presented here are generated on a  $300^3$  lattice.

Figure 1 shows the snapshots of the protease at the end of simulation for  $1 \times 10^7$  time steps. Small and large-scale globules clearly seen in this figure are just snapshots among numerous possible conformations. Visual inspection suggests that the hydrophobic residues constitute the core of these globule formations. The conformational phase space for such a protein is too large to pin down a unique structure. The global quantities such as the radius of gyration, energy, transport and mobility averaged over a large number of configurations capture their general characteristics. Let us look at some of the local quantities such as energy and mobility profiles.

The energy profile of each residue for the interaction strength  $f=150$  and 200 is presented in Fig. 2. The general pattern in residue energy appears to be preserved (with  $f=150, 200$ ) with somewhat lower energy among the lowest energy residues at higher interaction strength ( $f=200$ ). Least energetic residues are easy to identify which includes Arg( $R$ )<sup>41</sup> and Arg( $R$ )<sup>140</sup>, the residue with strongest electrostatic strength.<sup>30</sup> Many electrostatic and polar residues such as Glu( $E$ )<sup>21</sup>, Glu( $E$ )<sup>120</sup>, Asp( $D$ )<sup>25</sup>, Asp( $D$ )<sup>124</sup>, Tyr( $Y$ )<sup>59</sup>, Tyr( $Y$ )<sup>158</sup>, etc. have comparable energy. The symmetry in the energy of the residue in two monomers is remarkable, e.g.,  $P^1$ - $F^{99}$  of the first monomer and  $P^{100}$ - $F^{198}$  of the second such as residues  $R^{41}$  and  $R^{140}$ ,  $Y^{59}$  and  $Y^{158}$ , etc. Perhaps this is the cause for the symmetric structure for the protease with flaps<sup>1,6,9</sup> pointed out in the beginning. Based on the energy profiles one can identify the segmental stability or instability. For example, it is easier to find the least energy segment ( $D^{29}$ - $E^{35}$ ,  $D^{128}$ - $E^{134}$ ), the most energetic segments

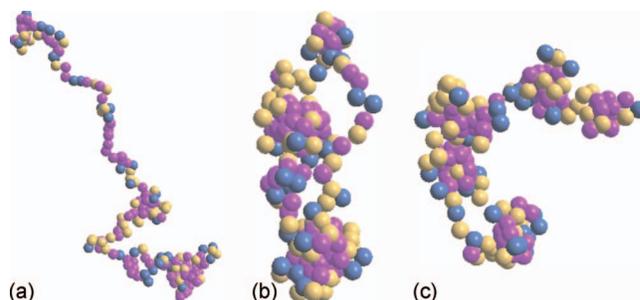


FIG. 1. [pink (hydrophobic, *H*), gold (polar, *P*), blue (electrostatic, *E*)] Snapshots of the 1DIFA dimer at the end of  $1 \times 10^7$  time steps with the interaction strength  $f=150, 200$ , and 250 on a  $300^3$  lattice.

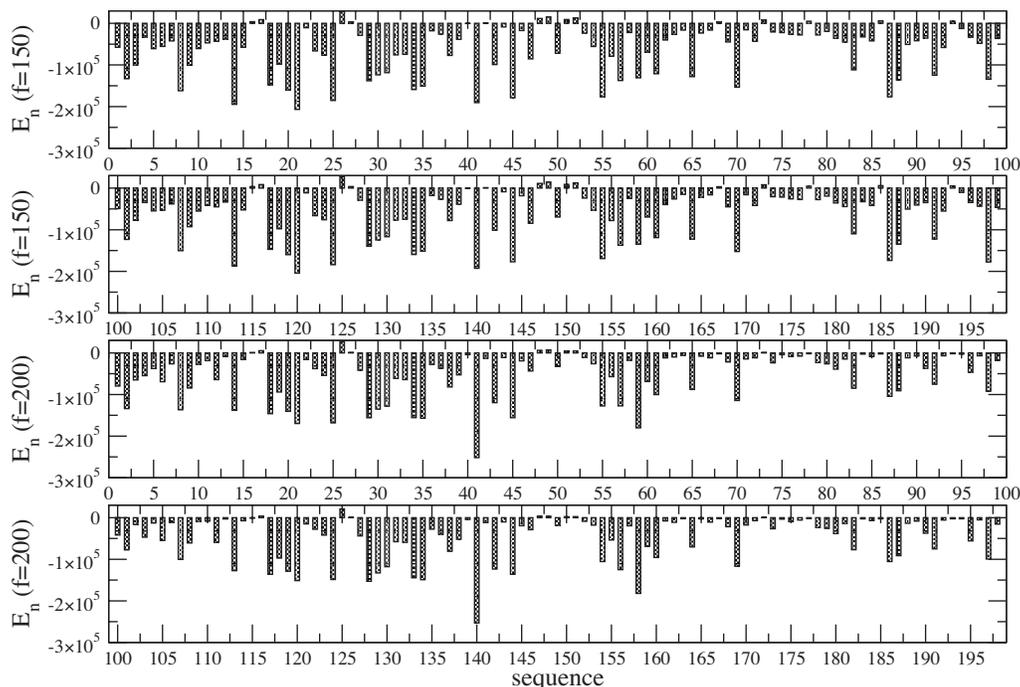


FIG. 2. Energy of each residue in HIV-1 protease (1DIFA dimer) with 198 residues; two monomers (1–99 and 100–198 in the sequence) are separated for clarity. Sample size  $300^3$  is used for time steps  $10^7$  with 50 independent runs for each interaction  $f=150$  and 200.

( $A^{71}$ - $V^{82}$ ,  $A^{170}$ - $V^{181}$ ), and so on. Thus, the energy profiles provide a useful map of the segmental energy and possibly their reactivity with regard to proteolysis.

The corresponding mobility profile of the residues is presented in Fig. 3. The mobility ( $M_n$ ) of each residue is defined as the average distance traveled per unit time step. Obviously the residues are more mobile with lower interaction strength ( $f=150$ ) than those with higher interaction strength ( $f=200$ ).

The symmetry in the mobility profiles of the residues in two monomers in 1DIFA dimer persists as with their energy (Fig. 2). The distinction among different segments of the protein based on residue mobility becomes more transparent at higher interaction strength ( $f=200$ ). The most mobile longest segment Asp<sup>25</sup>-Lys<sup>45</sup> in the first monomer and Asp<sup>124</sup>-Asp<sup>144</sup> in the second monomer consists of polar and electrostatic residues with relatively smaller numbers of hydrophobic residues (Gly, Ala, and Met). Some of these highly

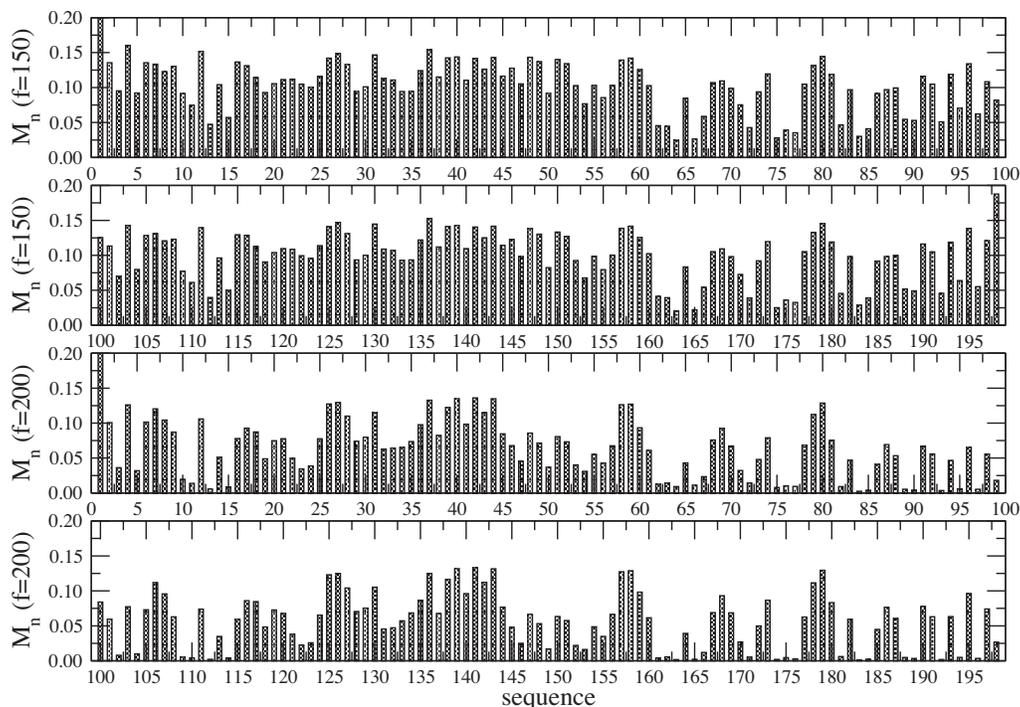


FIG. 3. Mobility of each residue in HIV-1 protease (1DIFA dimer) with 198 residues; two monomers are separated for clarity. Sample size  $300^3$  is used for time steps  $10^7$  with 50 independent runs for each interaction  $f=150$  and 200.

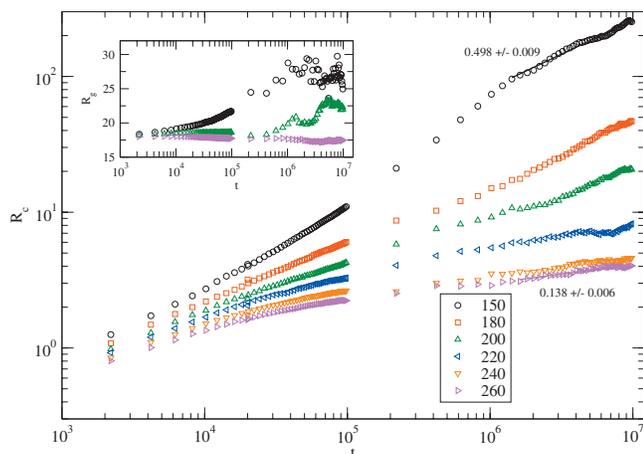


FIG. 4. Root mean square displacement of the center of mass of the protein (1DIFA dimer) with the time step for various interaction strength parameters  $f=150$ – $260$ ; the slopes of the long time data sets for two interaction strengths ( $f=150, 260$ ) are indicated. The inset shows the evolution of the radius of gyration ( $R_g$ ) with the time steps for representative values of  $f=150, 200$ , and  $260$ . Larger numbers (1000–2000) of independent samples were used for data sets with shorter time (up to  $10^5$ ) steps and lower numbers (50–80) of samples for long time data sets on a  $300^3$  lattice.

mobile residues are also the least energetic (Asp<sup>29</sup>-Glu<sup>35</sup>, Asp<sup>128</sup>-Glu<sup>134</sup>) which are interspersed between high mobile segments Thr<sup>26</sup>-Ala<sup>28</sup> in the first monomer and Thr<sup>125</sup>-Ala<sup>127</sup> in the second and Ser<sup>37</sup>-Pro<sup>44</sup> in the first monomer and Ser<sup>136</sup>-Pro<sup>143</sup> in the second. There are several highly mobile small segments separated by least mobile residues and segments. For example, in the first monomer, these high mobility segments are Trp<sup>6</sup>-Pro<sup>9</sup>, Thr<sup>12</sup>, Gly<sup>16</sup>-Gln<sup>18</sup>, Arg<sup>57</sup>-Gln<sup>61</sup>, and Gly<sup>78</sup>-Pro<sup>81</sup> which are separated by least mobile residues and segments Ile<sup>3</sup>, Leu<sup>5</sup>, Leu<sup>10</sup>, Val<sup>11</sup>, Ile<sup>13</sup>, Ile<sup>15</sup>, Ile<sup>62</sup>-Ile<sup>64</sup>, Ile<sup>66</sup>, Cys<sup>67</sup>, Ile<sup>72</sup>, Val<sup>75</sup>-Val<sup>77</sup>, Val<sup>82</sup>, Ile<sup>84</sup>, Ile<sup>85</sup>, Leu<sup>89</sup>, Leu<sup>90</sup>, Ile<sup>93</sup>, Cys<sup>95</sup>, and Leu<sup>97</sup>. The same residues are also in the second monomer. Note that all the least mobile residues are hydrophobic which seem to pin down the movements of neighboring residues. The mobility profile depends on the sequence and provides a road map of specificity of the enzyme. Despite symmetry in energy and mobility profiles in two monomers of the dimer, there are some asymmetries (e.g.,  $P^1$ ). Regardless of the remarkable symmetries, it is difficult to identify the configurations resembling a face with flaps.<sup>1,6,9</sup>

How does the mobility of individual residues collectively affect the global dynamics of the protein? Variation in the rms displacement ( $R_c$ ) of the center of mass of the protein with the time steps ( $t$ ) as presented in Fig. 4 for a range of interaction strength ( $f=150, 260$ ) can provide some insight into the global dynamics. The slope of  $R_c$  versus  $t$  on a log-log scale, i.e., the power-law exponent  $\nu$  in  $R_c \propto t^\nu$ , describes the characteristic dynamics. At a relatively low interaction strength ( $f=150$ ), the increase in  $R_c$  with the time step with a relatively high rate shows that the protein chain continues to diffuse ( $\nu \sim 1/2$ ) freely in the long time regime. This is consistent with the relatively high mobility profile of the residues (see Fig. 3 with  $f=150$ ). The dynamics of the protein slow down on increasing the interaction strength especially in the asymptotic regime after its relaxation. With

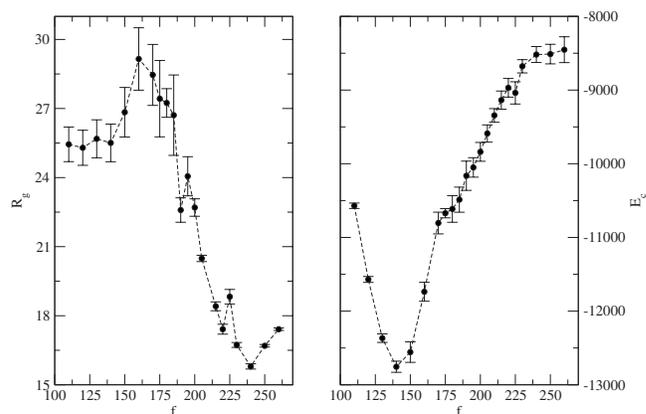


FIG. 5. Radius of gyration ( $R_g$ ) left and energy ( $E_c$ ) of 1DIFA dimer vs the interaction strength  $f$  in steady state with 50 independent samples on a  $300^3$  lattice.

the strong interaction strength ( $f \geq 220$ ), the global dynamics become very slow ( $\nu \approx 0.14$ ) as the hydrophobic residues pin down the protein chain; the onset of pinning is already seen in the mobility profile of the residues with  $f=200$  (Fig. 3). The quality of solvent, a measure of the interaction strength, thus dictates the global dynamics of the protein.

How does the global conformation of the protein relax starting from an initial random configuration? The inset in Fig. 4 shows the evolution of the radius of gyration ( $R_g$ ) of the protein with the time steps for representative values of the interaction strength, i.e.,  $f=150, 200$ , and  $260$ . Note that a million time steps are not long enough to reach a steady-state value with  $f=150, 200$  and therefore simulations had to be continued for much longer time ( $1 \times 10^7$ ) steps. The asymptotic value of  $R_g$  exhibits large fluctuations in the long time regime ( $t \approx 10^6$ – $10^7$  steps); the statistical fluctuations can be reduced by increasing the number of independent samples (compare short and long time data sets in the inset Fig. 4). The steady-state value of  $R_g$  is evaluated by averaging data in the asymptotic regime, i.e.,  $t \approx 10^6$ – $10^7$  for  $R_g$  with  $f=150$  and  $t \approx 5 \times 10^6$ – $10^7$  for  $R_g$  with  $f=200$ . Strong interactions ( $f=260$ ) pin down the conformations to their smaller values of  $R_g$ .

Variations in the steady-state gyration radius ( $R_g$ ) and energy ( $E_c$ ) of the protein with the interaction strength ( $f$ ) are presented in Fig. 5. Despite large error bars, both  $R_g$  and  $E_c$  show nonmonotonic dependence on the interaction strength with opposite trends. The smallest size of the protein ( $R_g$  in left figure) is not near the minimum energy (right figure) of the protein around  $f=130$ . In fact, the smallest value of the gyration radius appears around the interaction strength  $f=230$  where the protein energy is closer to its maximum value. Pinning of the conformations by the hydrophobic residues at such high interaction strengths seems to provide seed for the protein chain to collapse (see Figs. 1 and 3). Driving the protein toward its globular structure by increasing the interaction strength leads to higher energy of the protein, a somewhat contradictory observation to the general notion of funneling pathways in the complex energy landscape. These observations for both local and global physical quantities for the dimer protein are similar to what we have observed for

the monomer. Two monomers in the 1DIFA dimer simply help enhance the active sites (energy, mobility, and pinning).

Regardless of the remarkable symmetries in these characteristics of the dimer, we are unable to see the signal of the stable structure that resembles a face with flaps.<sup>1,6,9</sup> It is difficult to search all conformational phase spaces in such a large ensemble of the protein structure. The importance of sampling used in our MC simulation via the Metropolis algorithm provides an efficient searching procedure for the most accessible states (conformations) energetically. However, it does not capture all states. Therefore, it is not feasible to look at each configuration and identify the precise folded configuration explicitly as provided in the PDB. Apart from coarse graining the atomistic details the selective sampling though efficient and feasible may be one of the unavoidable limitations of the MC simulations. Such an approach is more valuable in identifying the trends in the ensemble averaged properties, e.g., correlation between the rapid decay of the radius of gyration with the interaction and the energy (Fig. 5) than identifying the specific structure. One of the primary implications of our study is that one should not rely on a single configuration to interpret experimental observables. Instead, one should consider ensemble averaged quantities for trends in observables including relaxation from segmental characteristics to global structure.

In conclusion, MC simulations are performed to examine the local and global properties of a HIV-1 protease, 1DIFA dimer consisting of 198 residues. A coarse-grained description is used to model the protease on a cubic lattice where empty lattice sites constitute the effective solvent medium. A residue is represented by a node of the protease chain and interacts with other residues within the chain and with the solvent sites within a range of interactions. The interaction strength is varied via a parameter  $f$  that represents the depth of the LJ potential well. Both local and global properties, conformation and dynamics depend on the interaction strength. The energy profile of the residues in the steady state reveals that the segments of proteins that consist of polar and electrostatic residues equilibrate to their lowest energy configuration. For example, some of the low energy residues are Glu<sup>21</sup>, Asp(D)<sup>25</sup>, Arg<sup>41</sup>, and Tyr(Y)<sup>59</sup> in first monomer and corresponding residues in the second monomer are Glu(E)<sup>120</sup>, Asp(D)<sup>124</sup>, Arg(R)<sup>140</sup>, and Tyr(Y)<sup>158</sup>. Hydrophobic residues (Ile, Leu, and Val) possess higher energy but are less mobile than their neighboring electrostatic and polar groups. Local segmental dynamics and structure are pinned down by these hydrophobic residues, which constitute the core of the globules at higher interaction strength. There is a remarkable symmetry in both energy and mobility profiles of the residues in each monomer of the 1DIFA dimer. Despite these symmetries in energy and mobility profiles of each monomer, it is difficult to identify the symmetric configurations resembling a face with flaps due to large phase space with enormous numbers of configurations.

Variation in the RMS displacement with the time steps

shows that the protein diffuses in a solvent with relatively weaker interaction. The global motion of the protein slows down systematically on increasing the interaction strength, where pinning by the hydrophobic core residues induce globule formation. Variations in the gyration radius ( $R_g$ ) and energy ( $E_c$ ) of the protein show non-monotonic dependence on the interaction strength with opposite trends. The smallest value of the gyration radius appears around the interaction strength ( $f=230$ ) where the protein energy is toward its maximum value. Pinning of the conformations by the hydrophobic residues at such high interaction strength seems to provide seed for the protein chain to collapse.

## ACKNOWLEDGMENTS

Support from the Materials and Manufacturing Directorate of the Air Force Research Laboratory is gratefully acknowledged. We thank Diana Flosenzer for careful reading and editing the manuscript.

- <sup>1</sup> See [http://en.wikipedia.org/wiki/Aspartic\\_acid\\_protease](http://en.wikipedia.org/wiki/Aspartic_acid_protease) for HIV-1 protease.
- <sup>2</sup> S. Piana, P. Carloni, and M. Parrinello, *J. Mol. Biol.* **319**, 567 (2002).
- <sup>3</sup> M. Cascella, C. Micheletti, U. Rothlisberger, and P. Carloni, *J. Am. Chem. Soc.* **127**, 3734 (2005).
- <sup>4</sup> M.A. Neri, M. Baaden, V. Carnevale, C. Anselmi, A. Maritan, and P. Carloni, *Biophys. J.* **94**, 71 (2008).
- <sup>5</sup> V. Carnevale, S. Raugei, C. Micheletti, and P. Carloni, *J. Phys. Chem.* **111**, 12332 (2007).
- <sup>6</sup> J. E. Foulkes-Murzycki, W. R. Scott, and C. A. Schiffer, *Structure* **15**, 225 (2007).
- <sup>7</sup> J. M. Louis, R. Ishima, I. Nesheiwat, L. K. Pannell, S. M. Lynch, D. A. Torchia, and A. M. Gronenborn, *J. Biol. Chem.* **278**, 6085 (2003).
- <sup>8</sup> M. Lapins, M. Eklund, O. Spjuth, P. Prusis, and J. E. S. Wkberg, *BMC Bioinf.* **9**, 181 (2008).
- <sup>9</sup> D. Li, M. S. Liu, B. Ji, K. Hwang, and Y. Huang, *J. Chem. Phys.* **130**, 215102 (2009).
- <sup>10</sup> Z. Bagci, A. Kloczkowski, R.L. Jernigan, and I. Bahar, *Proteins* **53**, 56 (2003).
- <sup>11</sup> H. Behringer, A. Degenhard, and F. Schmid, *Phys. Rev. Lett.* **97**, 128101 (2006).
- <sup>12</sup> P. G. Wolynes, J. N. Onuchic, and D. Thirumalai, *Science* **267**, 1619 (1995).
- <sup>13</sup> M. Skepö, P. Linse, and T. Arnebrant, *J. Phys. Chem. B* **110**, 12141 (2006).
- <sup>14</sup> B. Gillespie and K. W. Plaxco, *Annu. Rev. Biochem.* **73**, 837 (2004).
- <sup>15</sup> J. R. Banavar and A. Maritan, *Rev. Mod. Phys.* **75**, 23 (2003).
- <sup>16</sup> Y. Zhou and M. Karplus, *Nature (London)* **401**, 400 (1999).
- <sup>17</sup> M. Shen and K. F. Freed, *Proteins* **49**, 439 (2002).
- <sup>18</sup> B. Kuhlman, G. Dantas, G. C. Ireton, G. Varani, L. Barry, B. L. Stoddard, and D. Baker, *Science* **302**, 1364 (2003).
- <sup>19</sup> E. J. Sorin and V. S. Pande, *Biophys. J.* **88**, 2472 (2005).
- <sup>20</sup> Y. Wei, W. Nadler, and U. H. E. Hansmann, *J. Chem. Phys.* **128**, 025105 (2008).
- <sup>21</sup> T. Herges and W. Wenzel, *Phys. Rev. Lett.* **94**, 018101 (2005).
- <sup>22</sup> A. Liwo, M. R. Pincus, R. J. Wawak, S. Rackovsky, and H. A. Scheraga, *Protein Sci.* **2**, 1697 (1993).
- <sup>23</sup> J. Skolnick and A. Kolinsky, *J. Mol. Biol.* **221**, 499 (1991).
- <sup>24</sup> K. M. Fiebig and K. A. Dill, *J. Chem. Phys.* **98**, 3475 (1993).
- <sup>25</sup> B. Gertsman and Y. Garbourg, *J. Polym. Sci., Part B: Polym. Phys.* **36**, 2761 (1998).
- <sup>26</sup> P. P. Chapagain and B. Gertsman, *Biopolymers* **81**, 167 (2006).
- <sup>27</sup> C.-M. Chen and C.-C. Chen, *Biophys. J.* **84**, 1902 (2003).
- <sup>28</sup> J. Bjursell and R. B. Pandey, *Phys. Rev. E* **70**, 052904 (2004).
- <sup>29</sup> R. B. Pandey and B. L. Farmer, *Phys. Rev. E* **77**, 031902 (2008).
- <sup>30</sup> R. B. Pandey and B. L. Farmer, *J. Chem. Phys.* **130**, 044906 (2009).