

## Molecular Therapeutic Target for Type-2 Diabetes

Kuo-Chen Chou<sup>\*,†,‡</sup>

*Gordon Life Science Institute, San Diego, California 92130, and  
Tianjin Institute of Bioinformatics and Drug Discovery (TIBDD), Tianjin, China*

Received August 19, 2004

Many lines of evidences indicate that increased flux of glucose through the pathway, in which glutamine:fructose-6-phosphate amidotransferase (GFPT or GFAT) is a key catalyst while uridine-5'-diphosphate-*N*-acetylglucosamine (UDP-GlcNAc) functions as an energy sensor, can lead to the insulin resistance that is characteristic of Type-2 diabetes. In view of this, GFAT and its interaction mechanism with UDP-GlcNAc may become a novel therapeutic target for the treatment of type 2 diabetes. To stimulate the structure-based drug design, the three-dimensional structures of human GFAT1 monomer and dimer have been developed. It has been found by docking UDP-GlcNAc to the dimer (the smallest unit for catalyzing the substrate) that UDP-GlcNAc is bound to the interface of the dimer by 12 hydrogen bonds. On the basis of the docking results, a binding pocket of human GFAT1 dimer for UDP-GlcNAc is defined. All of these findings can serve as a reference or footing in developing new therapeutic strategy for the treatment of type-2 diabetes.

**Keywords:** diabetes • human GFAT • monomer • dimer • UDP-GlcNAc • docking • three-dimensional structure • binding pocket • hydrogen bonding • structural bioinformatics

### I. Introduction

There are three main types of diabetes: (1) Type-1 diabetes, also known as insulin-dependent diabetes mellitus (IDDM) or juvenile-onset diabetes. In this type of diabetes, the pancreas produces little or no insulin. (2) Type-2 diabetes, also known as noninsulin-dependent diabetes mellitus (NIDDM) or adult onset diabetes. People with this type of diabetes have a combination of insulin resistance and a defect in insulin production. (3) Gestational diabetes. This type of diabetes only affects pregnant women, and usually resolves once the pregnancy is over. However, women who experience gestational diabetes have a greater risk for developing type-2 diabetes later in life. Of these three forms of diabetes, type-2 accounts for 90–95% of all cases in the United States. People with type-2 diabetes usually develop the condition after age 45, and the risk for getting it increases with age. About 18% of the U. S. population 65 and older has type-2 diabetes. However, the number of children with type-2 diabetes is increasing rapidly. To develop effective treatment for diabetes, the first important thing is to understand the molecular mechanism that causes the disease.

Human beings need food to generate energy for growth and activities. The carbohydrates human beings consume are eventually broken down into a simple sugar called glucose, which passes into the bloodstream. For glucose to be used by cells in the body, a hormone produced by the pancreas, called insulin, is needed. For diabetic patients, the pancreas produces

little or no insulin, or the cells throughout the body malfunction in responding to the insulin that is being produced. The consequence is a build-up of glucose in the blood that eventually spills over into the urine. Elevated blood glucose levels are the “culprit” responsible for the many health problems associated with diabetes.

It was suggested via *in vitro* and *in vivo* studies that the increased flux of glucose through the hexosamine biosynthetic pathway may contribute to glucose-induced insulin resistance and to the induction of the synthesis of growth factor.<sup>1,2</sup> Because the first and rate-limiting step in the formation of hexosamine products is catalyzed by glutamine:fructose-6-phosphate amidotransferase (GFPT or GFAT), this enzyme is the key regulator in this pathway and is therefore also possibly involved in the alterations occurring in preclinical or manifest diabetic patients.<sup>2</sup> Overexpression of GFAT in the liver of transgenic mice results in enhanced glycogen storage, hyperlipidemia, obesity, and impaired glucose tolerance.<sup>3</sup> It is also known that GFAT catalyzes the conversion of fructose-6-phosphate (F-6-P) to glucosamine-6-phosphate (GlcN-6-P) with glutamine as an amino-donor,<sup>4</sup> and that GlcN-6-P is very rapidly further converted and activated to uridine-5'-diphosphate-*N*-acetylglucosamine (UDP-GlcNAc), serving as essential substrate for protein glycosylation.<sup>5</sup> Accordingly, finding the 3D (three-dimensional) structure of GFAT and its interaction mechanism with UDP-GlcNAc will provide us useful insights for developing new strategy for the treatment of type-2 diabetes.

Three human GFAT isoforms have been cloned, i.e., GFAT1, GFAT2, and GFAT1L.<sup>4,6–9</sup> GFAT2 is of less interest here because it appears to have highest expression in the brain.<sup>7</sup> GFAT1 is

\* To whom correspondence should be addressed. E-mail: kchou@san.rr.com.

† Gordon Life Science Institute.

‡ Tianjin Institute of Bioinformatics and Drug Discovery (TIBDD).

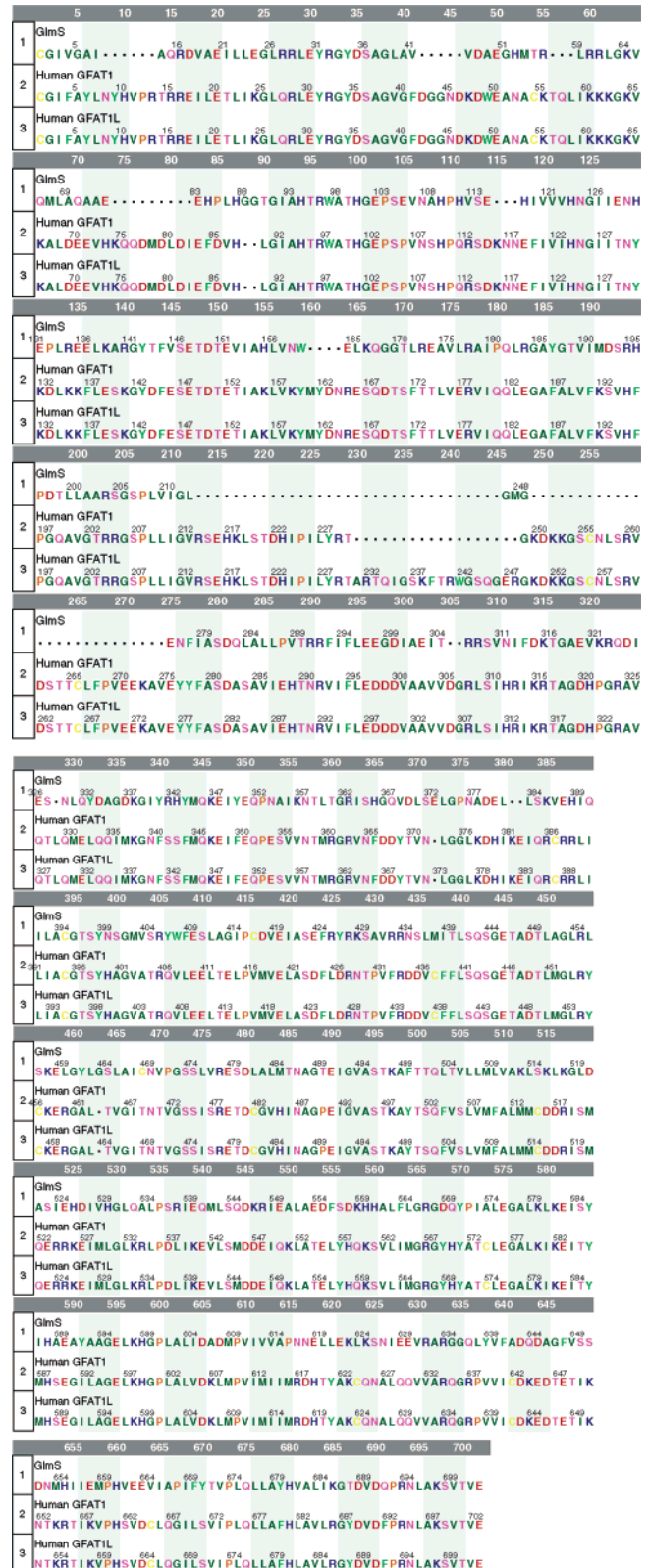
in liver and fat, and hence is an important target against diabetes and obesity. GFAT1L is a splice variant of GFAT1, and is mainly in muscle.<sup>10</sup> The only sequence difference between human GFAT1 and human GFAT1L is that GFAT1L contains an 18 amino acid segment insertion,<sup>10</sup> or a 54-bp insertion within the GFAT1 coding sequence.<sup>6</sup> Since structural bioinformatics has increasing impacts on biomedical science,<sup>11</sup> the present study was initiated in an attempt to develop the 3D structures of human GFAT1 monomer and dimer, and investigate their interactions with the ligands concerned, particularly UDP-GlcNAc, in hope to provide useful insights for stimulating new therapeutic strategy for the treatment of type-2 diabetes.

## II. Human GFAT1 Monomer

The sequence of human GFAT1 was taken from McKnight et al.,<sup>4</sup> and that of human GFAT1L taken from DeHaven et al.<sup>10</sup> The former contains 681 amino acid residues, and the latter contains 697 residues.

The template used for developing the 3D structure of GFAT1 was 1jax.pdb,<sup>12</sup> the crystal structure of *E.coli* GFAT (GlmS). The bacteria ortholog of GFAT has 608 residues. The sequence alignment of GlmS, human GFAT1, and human GFAT1L was performed by the PILEUP program in the GCG package.<sup>13</sup> The alignment result is shown in Figure 1, where the amino acids are colored according to their function: acidic – red; basic – blue; neutral hydrophilic – pink; aliphatic – dark green; aromatic – light green; thiol containing – yellow; and imino – orange. It was found during the alignment operation that the sequence similarity and identity between human GFAT1 and GlmS are 53.8% and 43.1%, respectively.

On the basis of the sequence alignment of Figure 1 as well as the template of GlmS (1jax.pdb), the atomic coordinates for GFAT1 monomer was derived by the segment matching approach.<sup>14–18</sup> The operation consisted of the following steps. (1) The target chain was first broken into short segments of sequence. (2) The database (formed by more than 5200 high-resolution crystal protein structures) was searched for matching segments according to the sequence alignment of Figure 1 and the shape of the template protein chain (1jax.pdb). (3) These segments coordinates were fitted into the growing target structure under the monitor to avoid any van der Waal overlap until all atomic coordinates of the target structure were obtained. (4) The process was repeated 10 times and an average model was generated, followed by energy minimization to create the final 3D structure. The segment matching approach was previously used to model the structure of the protease domain of caspase-8, at a time before the crystal coordinates were released for caspase-3.<sup>19</sup> In that particular study, the atomic coordinates of the catalytic domain of caspase-3 were predicted based on the crystal structure of caspase-1, and then the caspase-3 structure thus obtained served as a template to model the protease domain of caspase-8. After the crystal coordinates of caspase-3 protease domain were finally released and the crystal structure of the caspase-8 protease domain was determined,<sup>20</sup> it turned out that the RMSD (root-mean-square-deviation) for all the backbone atoms of the caspase-3 protease domain between the crystal and predicted structures was 2.7 Å, whereas the corresponding RMSD was 3.1 Å for caspase-8, and only 1.2 Å for its core structure. This indicates that the computed structures of caspase-3 and -8 were quite close to the corresponding crystal structures. Later on, this method was successively applied to model the CARDS (caspase recruitment domains) of Apaf-1, Ced-4, and Ced-3, based on the NMR



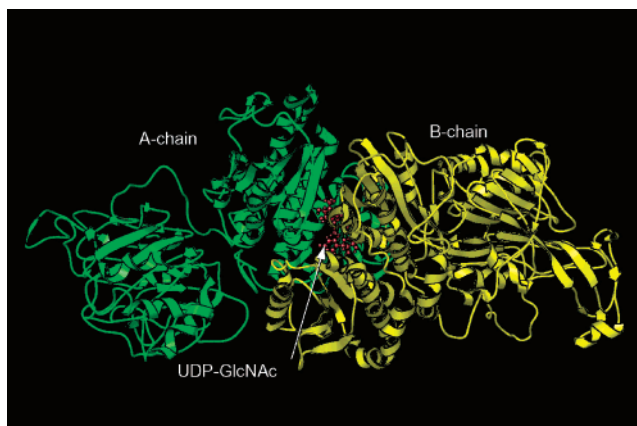
**Figure 1.** Sequence alignment of GlmS, human GFAT1, and human GFAT1L, where the amino acids are colored according to their function: acidic – red; basic – blue; neutral hydrophilic – pink; aliphatic – dark green; aromatic – light green; thiol containing – yellow; and imino – orange. The only difference of GFAT1L with GFAT1 is an 18 amino acid segment insertion at the position of residues 230–247. The sequence similarity and identity between GFAT1 and GlmS are 53.8% and 43.1%, respectively.



**Figure 2.** Computed 3D structure of human GFAT1 monomer, and its binding interaction with GLU and G6Q: the binding site for GLU is within N-domain of the enzyme, while that for G6Q is within C-domain.

structure of the RAIDD CARD,<sup>21</sup> and to model the Cdk5–Nck5a\* complex<sup>22</sup> as well as the protease domain of caspase-9 (23). It is intriguing to mention that, two years after the computed Cdk5–Nck5a\* complex structure was published,<sup>22</sup> the corresponding crystal structure was determined.<sup>24</sup> It has been found that the predicted Cdk5 and the crystal Cdk5 are almost the same. Furthermore, according to the report from the crystal structure, upon the binding of Cdk5 and Nck5a\* (or p25), the buried surface area is 3400 Å<sup>2</sup>,<sup>24</sup> which is quite close to 3461 Å<sup>2</sup> derived from the computed structure.<sup>11,22</sup> Also, based on the computed Cdk5–Nck5a\*–ATP structure, the molecular truncation experiments were conducted and it has been found that the experimental results “confirm and extend specific aspects of the original predicted computer model”.<sup>11,25</sup> Recently, it was also used to model the 3D structures of extracellular domains for the subtypes 1, 2, 3, and 5 of GABA-A receptors,<sup>26</sup> clarifying the ambiguity about the directionality of the subunit arrangement in the heteropentamers and providing useful insights for understanding the molecular operation mechanism of the receptors.

The overall structure thus obtained for the GFAT1 monomer is given in Figure 2. The binding sites for G6Q (Glucose-6-Phosphate) and glutamine are two important ligand binding sites of GFAT1, and they were located, respectively, as follows: (1) The binding site of human GFAT1 with G6Q can be easily derived because the crystal structure 1jax.pdb also contains the atomic coordinates of G6Q that is binding to GlnS. (2) Although 1jax.pdb does not contain the information for the ligand glutamine, the crystal structure of 1gdo.pdb, which is a complex of partial GlnS determined by Isupov et al.,<sup>27</sup> contains the atomic coordinates of glutamine and the catalytic domain of GlnS where the ligand is binding to. Thus, the binding site of human GFAT1 with glutamine was derived by superimposing the crystal structure (1gdo.pdb) onto the computed structure of GFAT1.



**Figure 3.** Computed 3D structure of human GFAT1 dimer, and its binding interaction with UDP–GlcNAc. The binding site is at the interface formed by the C-domains of the two constituent monomers (cf. Figure 2). See Table 1 for further details.

As shown in Figure 2, human GFAT1 monomer contains two domains: N-domain (res.1–310) and C-domain (res.311–680). The binding site of glutamine is within N-domain, while the binding site of G6Q within the C-domain.

### III. Human GFAT1 Dimer

Because of that UDP–GlcNAc is an important ligand to the hexosamine pathway, and that the smallest unit of GFAT1 with a fully competent catalytic capacity is a dimer, it is important to develop the 3D structure of human GFAT1 dimer as well. This can be done by performing the modeling procedure based on the same sequence alignment of Figure 1 but using both the B-chain and C-chain of the crystal structure of the 1jxa.pdb as a template. The detailed procedures are as follows: (1) Generate the 1st chain of human GFAT1 dimer using the B-chain of 1jxa.pdb as a template, with the presence of its C-chain as an environment. (2) Generate the 2nd chain of GFAT1 dimer using the C-chain of 1jxa.pdb as a template, with the presence of the new-generated 1st chain of GFAT1 dimer an environment. The presence of the other chain during the modeling process as described in the above steps is important in order to avoid the formation of any conflicting structure with overlap or penetration. The two chains thus generated were subject to an overall energy minimization with respect to all the atoms to finalize the entire structure for the dimer.

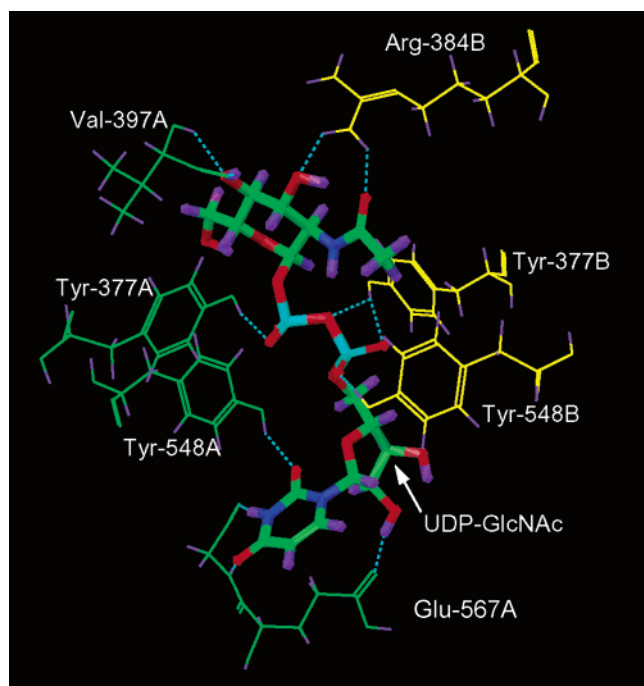
The 3D structure thus obtained for the human GFAT1 dimer is given in Figure 3, where A-chain is colored green and B-chain colored yellow. The interface of the dimer is formed by the C-domains (cf. Figure 2) of the two constituent monomers. While the crystal structures of *E.coli* GFAT (GlnS) in various complexes have been solved, this enzyme is not regulated by UDP–GlcNAc and hence the site to which UDP–GlcNAc binds on the human GFAT1 dimer could not be derived from the crystal structures of GlnS as done for the ligands G6Q and glutamine. To find out the binding site of UDP–GlcNAc, the computational docking was performed that is quite similar to the operation in finding the binding mechanism of coronavirus main proteinase with its ligands, such as KZ7088 and some relevant octapeptides.<sup>28,29</sup>

It was found that UDP–GlcNAc binds at the interface contacting residues on both polypeptide chains (Figure 3). These residues form a binding pocket for UDP–GlcNAc. The

**Table 1.** Residues<sup>a</sup> Involved in Forming the Binding Pocket of Human GFAT1 Dimer for UDP–GlcNAc

subunit A	subunit B
<b>Tyr-377<sup>b</sup></b>	<b>Tyr-377<sup>b</sup></b>
Val-381	His-378
Arg-384	Val-381
Met-396	<b>Arg-384<sup>b</sup></b>
Val-397	Val-397
Glu-398	Glu-398
	Leu-399
<b>Tyr-548<sup>b</sup></b>	<b>Tyr-548<sup>b</sup></b>
Leu-552	Leu-552
Ala-555	Ala-555
Lys-559	Lys-559
Ser-566	
<b>Glu-567<sup>b</sup></b>	Glu-567
<b>Gly-568<sup>b</sup></b>	Gly-568
Ile-569	Ile-569

<sup>a</sup> Residues listed here have at least one heavy atom with a distance  $\leq 5 \text{ \AA}^{22}$  from a heavy atom of UDP–GlcNAc. <sup>b</sup> Residue in bold-face type forms at least one hydrogen bond with UDP–GlcNAc.



**Figure 4.** Illustration to show the hydrogen bonding (green dotted line) interaction of human GFAT1 dimer with UDP–GlcNAc. Only 10 hydrogen bonds are shown. The hydrogen bond between O of Gly-568A and OH<sup>33</sup> of GlcNAc, and that between NH of Gly-568A and O<sup>40</sup> of GlcNAc are not shown in this view.

constituents of the pocket are defined by those residues that have at least one heavy atom, i.e., an atom other than hydrogen, with a distance  $\leq 5 \text{ \AA}^{22}$  from a heavy atom of UDP–GlcNAc. The pocket thus defined consists of 28 residues, of which 14 are from subunit A and 14 from subunit B. The detailed residues involved in forming the binding pocket are listed in Table 1. Furthermore, it has been observed that UDP–GlcNAc is bound to the dimer interface by 12 hydrogen bonds, of which 7 bonds are formed with Tyr-377, Val-397, Tyr-548, Glu-567, and Gly-568 of subunit A, and 5 formed with Tyr-377, Arg-384, and Tyr-548 of subunit B (see Table 1 and Figure 4).

## IV. Conclusion

Insulin resistance is defined as the inability of insulin to stimulate normal rates of glucose uptake and glycogen synthesis. The hexosamine pathway for UDP–GlcNAc biosynthesis in mammals functions as an energy sensor. Increased flux of glucose through this pathway can lead to the insulin resistance that is characteristic of type-2 diabetes. Human GFAT1 is in liver and fat that catalyzes the rate-limiting reaction in UDP–GlcNAc biosynthesis, and is a novel therapeutic target against type-2 diabetes.

On the basis of the crystal structure of *E.coli* GFAT (GlmS), the 3D structures of human GFAT1 monomer and dimer were developed. The binding sites of human GFAT1 with glutamine and G6P were derived, respectively. Because of that GlmS is not regulated by UDP–GlcNAc and that the dimer is the smallest unit for catalyzing UDP–GlcNAc, the computational docking was conducted to predict its binding site on a human GFAT1 dimer. It was found that UDP–GlcNAc is binding to the interface of the dimer. On the basis of the docking results, the binding pocket for UDP–GlcNAc has been defined that involves 12 residues from subunit A and 12 from subunit B. It has been observed that UDP–GlcNAc forms 7 hydrogen bonds with subunit A and 5 hydrogen bonds with subunit B. These findings can be used to guide mutagenesis studies and steer medicinal chemistry for synthesizing selective inhibitors for therapeutic treatment of type-2 diabetes.

**Abbreviations Used:** 3D, three-dimensional; GFAT (or GFPT), Glutamine:Fructose-6-Phosphate Amidotransferase; UDP–GlcNAc, uridine-5'-diphosphate-*N*-acetylglucosamine; G6P, Glucose-6-Phosphate

## References

- McClain, D. A.; Crook, E. D. *Diabetes* **1996**, *45*, 1003–1009.
- Nerlich, A. G.; Sauer, U.; Kolm-Litty, V.; Wagner, E.; Koch, M.; Schleicher, E. D. *Diabetes* **1998**, *47*, 170–178.
- Veerababu, G.; Tang, J.; Hoffman, R. T.; Daniels, M. C.; Hebert, L. F., Jr.; Crook, E. D.; Cooksey, R. C.; McClain, D. A. *Diabetes* **2000**, *49*, 2070–2078.
- McKnight, G. L.; Mudri, S. L.; Mathewes, S. L.; Traxinger, R. R.; Marshall, S.; Sheppard, P. O.; O'Hara, P. J. *J. Biol. Chem.* **1992**, *267*, 25208–25212.
- Kreppel, L. K.; Blomberg, M. A.; Hart, G. W. *J. Biol. Chem.* **1997**, *272*, 9308–9315.
- Niimi, M.; Ogawara, T.; Yamashita, T.; Yamamoto, Y.; Ueyama, A.; Kambe, T.; Okamoto, T.; Ban, T.; Tamanoi, H.; Ozaki, K.; Fujiwara, T.; Fukui, H.; Takahashi, E.; Kyushiki, H.; Tanigami, A. *J. Hum. Genet.* **2001**, *46*, 566–571.
- Oki, T.; Yamazaki, K.; Kuromitsu, J.; Okada, M.; Tanaka, I. *Genomics* **1999**, *57*, 227–234.
- Whitmore, T. E.; Mudri, S. L.; McKnight, G. L. *Genomics* **1995**, *26*, 422–423.
- Zhou, J.; Neidigh, J. L.; Espinosa, R.; LeBeau, M. M.; McClain, D. A. *Hum. Genet.* **1995**, *96*, 99–101.
- DeHaven, J. E.; Robinson, K. A.; Nelson, B. A.; Buse, M. G. *Diabetes* **2001**, *50*, 2419–2424.
- Chou, K. C. *Curr. Med. Chem.* **2004**, *11*, 2105–2134.
- TePLYakov, A.; Obmolova, G.; Badet, B.; Badet-Denisot, M. A. *J. Mol. Biol.* **2001**, *313*, 1093–1102.
- Devereux, J. 1994 in Genetic Computer Group (GCG), Madison, Wisconsin.
- Jones, T. A.; Thirup, S. *EMBO J.* **1986**, *5*, 819–822.
- Blundell, T. L.; Sibanda, B. L.; Sternberg, M. J. E.; Thornton, J. M. *Nature (London)* **1987**, *326*, 347–352.
- Finzel, B. C.; Kimatian, S.; Ohlendorf, D. H.; Wendoloski, J. J.; Levitt, M.; Salemme, F. R. In *Crystallographic and Modeling Methods in Molecular Design*; Bugg, C. E., Ealick, S. E., Eds.; Springer-Verlag: Berlin 1989 pp 175–188.
- Levitt, M. *J. Mol. Biol.* **1992**, *226*, 507–533.
- Chou, K. C.; Nemethy, G.; Pottle, M.; Scheraga, H. A. *J. Mol. Biol.* **1989**, *205*, 241–249.

- (19) Chou, K. C.; Jones, D.; Henrikson, R. L. *FEBS Lett.* **1997**, *419*, 49–54.
- (20) Watt, W.; Koeplinger, K. A.; Mildner, A. M.; Henrikson, R. L.; Tomasselli, A. G.; Watenpaugh, K. D. *Structure* **1999**, *7*, 1135–1143.
- (21) Chou, J. J.; Matsuo, H.; Duan, H.; Wagner, G. *Cell* **1998**, *94*, 171–180.
- (22) Chou, K. C.; Watenpaugh, K. D.; Henrikson, R. L. *Biochem. Biophys. Res. Commun.* **1999**, *259*, 420–428.
- (23) Chou, K. C.; Tomasselli, A. G.; Henrikson, R. L. *FEBS Lett.* **2000**, *470*, 249–256.
- (24) Tarricone, C.; Dhavan, R.; Peng, J.; Areces, L. B.; Tsai, L. H.; Musacchio, A. *Mol. Cell* **2001**, *8*, 657–669.
- (25) Zhang, J.; Luan, C. H.; Chou, K. C.; Johnson, G. V. W. *Proteins: Struct. Funct. Genet.* **2002**, *48*, 447–453.
- (26) Chou, K. C. *Biochem. Biophys. Res. Commun.* **2004**, *316*, 636–642.
- (27) Isupov, M. N.; Obmolova, G.; Butterworth, S.; Badet-Denisot, M. A.; Badet, B.; Polikarpov, I.; Littlechild, J. A.; Teplyakov, A. *Structure* **1996**, *4*, 801–810.
- (28) Chou, K. C.; Wei, D. Q.; Zhong, W. Z. *Biochem. Biophys. Res. Commun.* **2003**, *308*, 148–151 (Erratum: *ibid.*, **2003**, *310*, 675).
- (29) Sirois, S.; Wei, D. Q.; Du, Q.; Chou, K. C. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1111–1122.

PR049849V