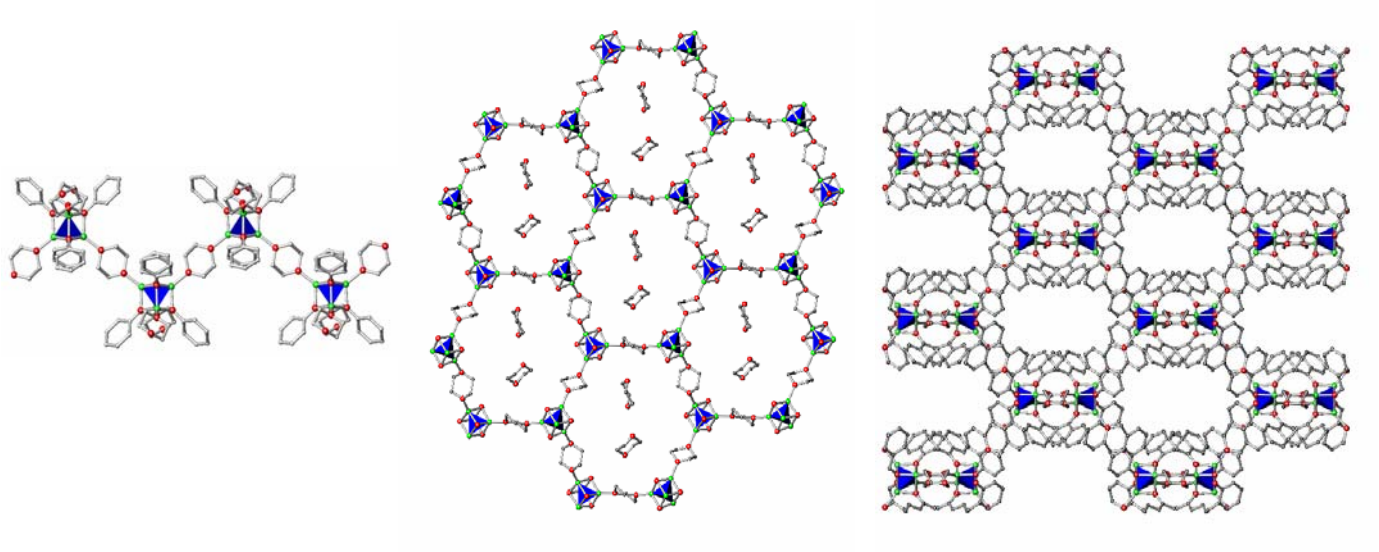
*Proceedings from the Symposium*

# Margaret C. Etter Early Career Award



## ETTER TRANSACTIONS OF THE AMERICAN CRYSTALLOGRAPHIC ASSOCIATION

Edited by Frances Jurnak

VOLUME 1, 2005

*Proceedings from the Symposium*

# Margaret C. Etter Early Career Award

ACA ETTER TRANSACTIONS
VOLUME 1

Walt Disney World, Orlando, Florida
May 28 – June 2, 2005

**Editor**

Frances Jurnak
University of California, Irvine
Tel: 949-824-6623
Fax: 949-824-8540
jurnak@uci.edu

# BIS-CHELATION AND ANION EFFECTS INVOLVING A MOLECULE CONSTRUCTED IN THE ORGANIC SOLID STATE USING MOLECULAR TEMPLATES

Tamara D. Hamilton, Giannis S. Papaefstathiou and Leonard R. MacGillivray
*University of Iowa Department of Chemistry, Iowa City, IA 52242*

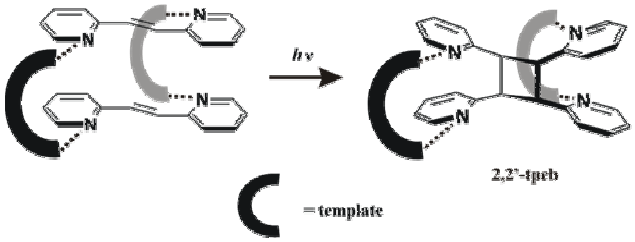
## 1. ABSTRACT

A molecule derived from a template-directed solid-state synthesis, *rctt*-tetrakis(2-pyridyl)cyclobutane (2,2'-tpcb), serves as a bis-chelating ligand upon reaction with $Cu(BF_4)_2{\cdot}H_2O$ and $NEt_4PF_6$ to produce the dinuclear complex $[Cu_2(OH)_2(H_2O)_4(\mu\text{-}2,2'\text{-tpcb})][PF_6]_2$ (**1**).

## 2. INTRODUCTION

Transition-metal-ion complexes involving seven-membered chelation rings have emerged as important for applications in catalysis and biology. Specifically, N-$C_4$-N bridged metallocycles [1] have found applications as catalysts [2] and models of biological compounds [3], as well as antibacterial [4] and antitumor [5] agents. In terms of catalysis, the application of chelating N-donor ligands - as compared to analogous polyphosphines - has begun to attract attention, especially for the activation of C-H bonds [6]. Thus, ligands that provide more than one pyridyl group in close proximity are being studied as preorganized metal chelation units [6b]. In addition, there is increasing interest in the development of novel non-metallocene catalysts for olefin polymerization for which complexes involving N-$C_4$-N bridged chelate rings have been effective [2c]. Therefore, studies that involve N-$C_4$-N ligands with a potential to offer seven-membered chelation are of much current interest.

The ability to position molecules in an orientation appropriate for reaction in the organic solid state using molecular templates that operate *via* hydrogen bonds has led to the quantitative and gram-scale construction of molecules difficult to obtain from solution [7,8]. In particular, we have shown that templates based on resorcinol can orient olefins [*e.g. trans*-1,2-bis(2-pyridyl)ethylene) (2,2'-bpe)] in the solid state in positions suitable for intermolecular [2+2] photodimerizations. The suitability of the products as ligands for coordination chemistry is evidenced by the presence of pyridine rings that emanate from central cyclobutane rings.




**Scheme 1.** Template-directed solid-state synthesis of 2,2'-tpcb from 2,2'-bpe.

As part of an ongoing study to elucidate anion effects involving transition-metal-ion complexes of *rctt*-tetrakis(2-pyridyl)cyclobutane (2,2'-tpcb), we wish to report here the synthesis and structure determination of the dinuclear Cu(II) complex $[Cu_2(OH)_2(H_2O)_4(\mu\text{-}2,2'\text{-tpcb})][PF_6]_2$ (**1**). In line with our previous study [7], the tetrapyridine acts as a bis-chelating bridge with two Cu(II) centers, giving rise to two seven-membered metallocycles. In contrast to our previous work, the $PF_6^-$ counter ions are non-coordinating and, thus, provide an opportunity for each Cu(II) ion to interact with additional donor units. In the case of **1**, each Cu(II) center is coordinated by two water molecules and a single hydroxide anion, each of which participates in hydrogen bonds with the $PF_6^-$ anion.

## 3. METHODS

Complex **1** was prepared by combining methanolic solutions of 2,2'-tpcb (0.025 g, 0.068 mmol) [8a] and $Cu(BF_4)_2{\cdot}H_2O$ (0.032 g, 0.137 mmol) (1:2 ratio). To the resulting clear blue solution was added a methanolic solution of $NEt_4PF_6$ (0.037 g, 0.137 mmol) (25 mL methanol total). A 5 mL aliquot of the latter solution was then

layered with 10 mL benzene in a test tube and allowed to sit for a period of approximately one week. The solution produced blue single crystals (yield: 0.015 g, 29 %) suitable for X-ray analysis [9].

## 4. RESULTS AND DISCUSSION

Compound **1** crystallizes in the space group $P2_1/n$. An ORTEP perspective of the asymmetric unit of **1** is shown in **Figure 1a**. The asymmetric unit consists of one Cu(II) ion, one-half molecule of 2,2'-tpcb, two water molecules, one hydroxide ion, and one $PF_6^-$. As shown in **Figure 1b**, the dinuclear complex, which sits around a crystallographic center of inversion, contains one molecule of 2,2'-tpcb that bridges two Cu(II) centers [d(Cu1···Cu1a) 7.47 Å]. The Cu(II) ions of the complex adopt an *anti* conformation and exhibit a bite angle of 88.8° [10]. Similar to $[Cu_2(NO_3)_4(\mu\text{-}2,2'\text{-tpcb})]$ and $[Cu_2(\mu_2\text{-}SO_4)_2(\mu\text{ -}2,2'\text{-tpcb})(H_2O)_2]_\infty$, each seven-membered chelation ring adopts a boat conformation. The boat conformation also compares favorably to the chelation complexes of *bis*(2-pyridyl)ethane with Pt(II) [11,12] and 1,2-bis(6-methylpyridin-2-yl)ethane with Pd(II) [13]. The coordination geometry around each Cu(II) center of **1** is square pyramidal. Two pyridyl units of 2,2'-tpcb and two water molecules (O1 and O2) form the base of the pyramid while a hydroxide ion (O3) occupies the apical position. The metal-ligand distances around the base of the pyramid range from 1.96-2.02 Å, while the Jahn-Teller distorted apical distance is 2.21 Å (Table 1).

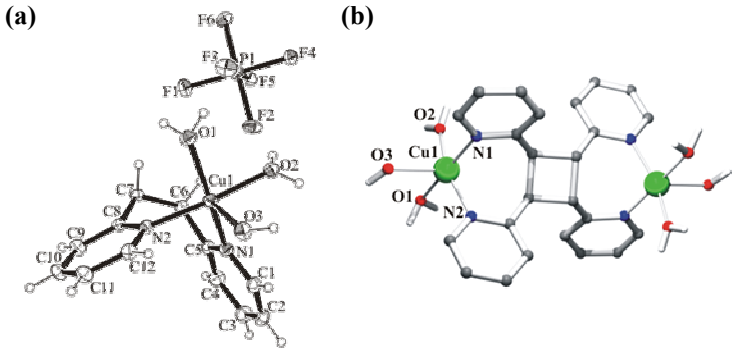



**Figure 1**. **(a)** ORTEP representation of the asymmetric unit of **1**. Displacement ellipsoids are drawn at the 30% probability level, with hydrogen atoms being assigned an arbitrary radius; **(b)** The dinuclear complex **1** (color scheme: green = copper, red = oxygen, blue = nitrogen, grey = carbon). Hydrogen atoms of 2,2'-tpcb have been omitted.

Views of the extended structure of **1** are shown in **Figures 2** and **3**. The complexes have assembled to form layers, parallel to the *ac*-plane, which exhibit an *abab* stacking pattern. The layers are held together by face-to-face π-π forces involving the ligated pyridyl groups (**Figure 2**). The counter $PF_6^-$ ions form layers sandwiched between the layers of Cu(II) complexes. The anions participate in extensive O-H···F hydrogen bonds that involve the two water molecules (O1, O2) and hydroxide ion (O3). Specifically, each $PF_6^-$ anion participates in a total of six O-H···F hydrogen bonds (**Figure 3a**) [d(O1···F3) 2.750(5) Å; d(O1···F4) 2.626(5) Å; d(O2···F2) 2.587(5) Å; d(O2···F6) 2.600(5) Å; d(O3···F1) 2.869(5) Å; d(O3···F5) 2.665(5) Å] (**Figure 3b**) with the ligands.

**TABLE 1**

Selected Bond Lengths (Å) and Angles (°) for **1**

| | | | |
|---|---|---|---|
| Cu1-O1 | 1.987(4) | O1-Cu1-O2 | 84.7(2) |
| Cu1-O2 | 1.963(4) | O1-Cu1-O3 | 99.7(2) |
| Cu1-O3 | 2.214(4) | O1-Cu1-N2 | 97.5(1) |
| Cu1-N1 | 2.020(4) | O2-Cu1-O3 | 83.7(2) |
| Cu1-N2 | 1.986(4) | O2-Cu1-N1 | 92.8(2) |
| | | O3-Cu1-N1 | 101.7(2) |
| | | O3-Cu1-N2 | 100.5(1) |
| | | N1-Cu1-N2 | 88.8(1) |

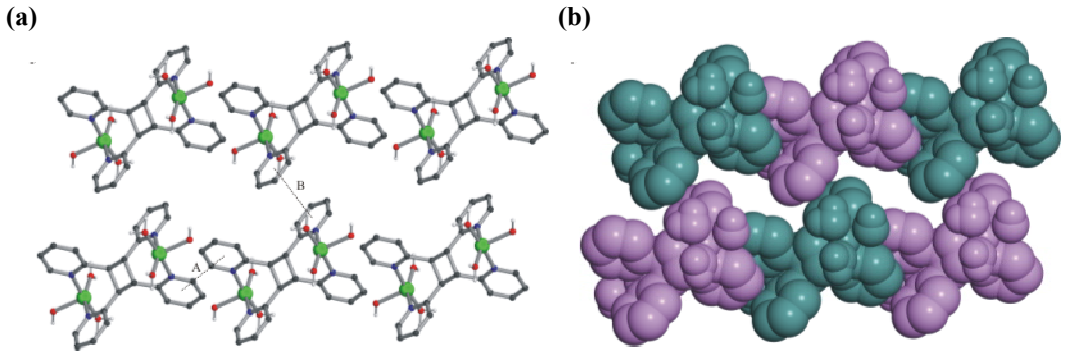



**Figure 2.** Extended structure of **1** showing the π-stacking of pyridine rings viewed in the *ac* plane: **(a)** ball-and-stick representation ($d_A$ = 3.697(5); $d_B$ = 3.874(5) Å) (color scheme: green = copper, red = oxygen, blue = nitrogen, grey = carbon) and **(b)** space-filling representation (hydrogen atoms of 2,2'-tpcb omitted for clarity).

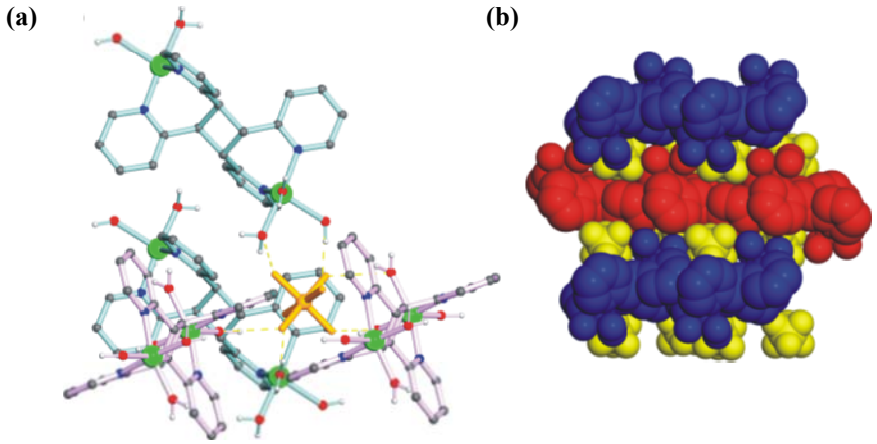



**Figure 3.** **(a)** Hydrogen bonding array involving $PF_6^-$ anion (gold) in **1**. Each F-atom is involved in one hydrogen bond (yellow) to a coordinated water or hydroxide molecule of a neighboring assembly (color scheme: green = copper, red = oxygen, blue = nitrogen, grey = carbon, white = hydrogen) (hydrogen atoms of 2,2'-tpcb omitted for clarity) and **(b)** space-filling representation of **1**, showing *abab* packing of complex layers alternating with $PF_6^-$ anions. Complexes are shown in red or blue, while $PF_6^-$ anions are shown in yellow.

## 5. CONCLUSION

In conclusion, template-directed solid-state organic synthesis has led to the construction of a tetrapyridyl ligand that forms the dinuclear coordination complex **1** with Cu(II) ions. The complex incorporates two seven-membered chelation rings [7]. 2,2'-tpcb is the first ligand to sustain seven-membered chelation in a series of coordination complexes. The non-coordinating behavior of the $PF_6^-$ counter-ion provides open sites on each metal, which are occupied by water molecules and hydroxide ions.

## 6. REFERENCES

[1] a) McCullough, F. Jr.; Bailar, J. C. Jr. *J. Am. Chem. Soc.* **1956**, *78*, 714-716;
b) Habu, T.; Bailar, J. C. Jr. *J. Am. Chem. Soc.* **1966**, *88*, 1128-1130.

[2] a) Gibson, V. C.; Kimberley, B. S.; White, A. J. P.; Williams, D. J.; Howard, P. *Chem. Commun.* **1998**, 313-314;
b) Kretzschmar, E. A.; Kipke, J.; Sundermeyer, J. *Chem. Commun.* **1999**, 2381-2382;
c) Siemeling, U.; Kolling, L.; Stammler, A.; Stammler, H. G.; Kaminski, E.; Fink, G., *Chem. Commun.* **2000**, 1177-1178;
d) Bacchi, A.; Pelizzi, G.; Minardi, G.; Pistuddi, A. M.; Solinas, C.; Chelucci, G. *Trans. Met. Chem.* **2002**, *27*, 274-278;
e) Minardi, G.; Mura, E.; Pistuddi, A. M.; Solinas, C.; Bacchi, A.; Pelizzi, C.; Pelizzi, G.; Chelucci, G *Trans. Met. Chem.* **1999**, *24*, 481-485;

f) Zhu, H. B.; Dai, Z.-Y.; Huang, W.; Cui, K.; Gou, S.-H.; Zhu, C.-J. *Polyhedron* **2004**, *23*, 1131-1137;
g) Bandoli, G.; Gerber, T. I. A.; Jacobs, R.; du Preez, J. G. H. *Inorg. Chem.*, **1994**, *33*, 178-179.

[3] a) Martin, J. W. L.; Timmons, J. H.; Martell, A. E.; Rudolf, P.; Clearfield, A. *Inorg. Chem.* **1981**, *20*, 814-821;
b) Choo P. L.; Mulichak, A. M.; Jones Jr., R. W.; Bacon, J. W.; Pett, V. B.; Zacharias, D. E. *Inorg. Chim. Acta* **1990**, *171*, 183-192;
c) Ball, R. G.; Trotter, J. *Can. J. Chem.* **1979**, *57*, 1368-1371;
d) McFarland, C. A.; Gross, S. A.; Winfield, J. S.; Pett, V. B. *Inorg. Chim. Acta*, **1994**, *221*, 35-41;
e) Balamurugan, R.; Palaniandavar, M.; Gopalan, R. S. *Inorg. Chem.* **2001**, *40*, 2246-2255.

[4] Nandi, A. K.; Chaudhuri, S.; Mazumdar, S. K. Ghosh, S. *Inorg. Chim. Acta* **1984** *92*, 235-240.

[5] a) Khokar, A. R.; Shamsuddin, S.; Xu, Q. *Inorg. Chim. Acta* **1994**, *219*, 193-197;
b) Shamsuddin, S.; van Haal, J. W.; Stark, J. L.; Whitmore, K. H.; Khokhar, A. R. *Inorg. Chem.* **1997**, *36*, 5969-5971;
c) Shamsuddin, S.; Santillan, C. C.; Stark, J. L.; Whitmire, K. H.; Siddik, Z. H.; Khokhar, A. R. *J. Inorg. Biochem.* **1998**, *71*, 29-35;
d) Hoeschele, J. D.; Showalter, H. D. H.; Kraker, A. J.; Elliott, W. L.; Roberts, B. J.; Kampf, J. W. *J. Med. Chem.* **1994**, *37*, 2630-2636.

[6] a) Vedernikov, A. N.; Huffman, J. C.; Caulton, K. G. *Inorg. Chem.* **2002**, *41*, 6244-6248;
b) Vedernikov, A. N.; Huffman, J. C.; Caulton, K. G. *Inorg. Chem.* **2002**, *41*, 6867-6874; Vedernikov, A. N.; Pink, M.; Caulton, K. G. *Inorg. Chem.* **2004**, *43*, 3642-3646.

[7] Hamilton, T. D.; Papaefstathiou, G. S.; MacGillivray, L. R. *Cryst. Eng. Comm.* **2002**, *4*, 223-226.

[8] a) Papaefstathiou, G. S.; Kipp, A. J.; MacGillivray, L. R. *Chem. Commun.* **2001**, 2462-2463;
b) MacGillivray, L. R.; Reid, J. L.; Ripmeester, J. A. *J. Am. Chem. Soc.* **2000**, *122*, 7817-7818;
c) Hamilton, T. D.; Papaefstathiou, G. S.; MacGillivray, L. R. *J. Am. Chem. Soc.* **2002**, *124*, 11606-11607.;
d) Varshney, D. B.; Papaefstathiou, G. S.; MacGillivray, L. R. *Chem. Commun.* **2002**, 1964-1965;
e) Gao, X.; Friščić, T.; MacGillivray, L. R. *Angew. Chem. Int. Ed.* **2004**, *43*, 232-236.

[9] Crystal data for **1**: monoclinic, $P2_1/n$, Z = 4, a = 9.716(5), b = 16.586(5), c = 9.939(5) Å, $\beta$ = 98.270(5) º, V = 1585.0(12) Å$^3$, 2738 unique reflections, $R$ = 0.0492 [for 2122 reflections with $I \geq 2\sigma(I)$], $R$ = 0.0742 and w$R^2$ = 0.1408 (all data). Crystallographic Data has been deposited with the Cambridge Crystallographic Data Centre and has been allocated with deposition number CCDC 291578.

[10] The bite angle is the angle between chelating ligand donor atoms (L-M-L) in a metal complex.

[11] McFarlane, A.; Lusty, J. R.; Fiol, J. J.; Terron, A.; Molins, E.; Miravitlles, C.; Moreno, V. *Z. Naturforsch., Teil B*, **1994**, *49*, 844-848.

[12] a) Marcelis, A. T. M.; Korte, H.-J.; Krebs, B.; Reedijk, J. *Inorg. Chem.*, **1982**, *21*, 4059-4063;
b) Kleibohmer, W.; Krebs, B.; Mercelis, A. T. M.; Reedijk, J.; Van der Veer, J. L. *Inorg. Chim. Acta* **1983**, *75*, 45-50.

[13] Baker, A. T.; Crass, J. K.; Kok, G. B.; Orbell, J. D.; Yuriev, E. *Inorg. Chim. Acta* **1993**, *214*, 169-176.

# USE OF ALKALI METAL AGGREGATES IN CONTROLLING NETWORK ASSEMBLY

Dugald J. MacDougall, J. Jacob Morris, Bruce C. Noll and Kenneth W. Henderson*
*Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, Indiana 46556, USA.*

## 1. ABSTRACT

This paper summarizes the first systematic studies into the use of *s*-block aggregates to control the rational assembly of network polymers. Preliminary work using lithium and sodium aryloxide aggregates has established the efficacy of this methodology, and in particular that the control of network architectures can be achieved on a regular basis. These studies also demonstrate that air-stable, robust, open-framework materials may be prepared through use of these systems. Overall, the paper outlines an area that has considerable potential for the preparation of new materials, that advance knowledge in the understanding of the self-assembling process.

## 2. INTRODUCTION

The synthesis and characterization of periodic network assemblies is an area of increasing international academic and industrial interest [1-4]. The attention paid to this topic can be attributed, in part, to the fundamental scientific challenges present in developing rational routes to structurally well-defined solids and also as a consequence of the potential utility of these materials in applications as diverse as catalysis [5], chemical separation [6], optics [7] and electronics [8]. In this regard, two main strategies have been adopted in the formation of frameworks from molecular precursors [9], firstly through the use of hydrogen-bonded organic solids [10] and secondly using metal-organic coordination polymers containing transition elements and ligand spacer molecules [11]. These strategies have also been combined where metal-containing complexes are incorporated within hydrogen-bonded networks [12]. In addition, a notable area of success in controlling the formation of covalently-linked solids has been the use of metal carboxylate clusters as directional components of metal-organic frameworks (MOFs) [13]. However, a prominent omission from the current list of strategies directed towards rational network synthesis is the use of early main group metals to dictate supramolecular structure. This deficit may be explained by the considerable potential problems associated with these species, such as the presence of complex dynamic solution equilibria between aggregated complexes; the spherical non-polarizable nature of the cations; the limited preference for fixed metal geometries, and the lability of the metal-ligand bonding. All of these issues may lead to poor predictability and reproducibility in network synthesis [14-16]. In this context it is unsurprising that researchers have opted for alternative structural scaffolds, and to a large extent have consciously avoided the use of *s*-block metal cations. In this paper we directly address these established perceptions and contest that the early main group elements can be successfully employed to control network assembly by the informed selection of metal, anionic partner, linker molecule and synthetic conditions. More specifically, we outline a novel approach to rational network synthesis utilizing pre-assembled *s*-block molecular aggregates as secondary building units (SBUs) to control supramolecular structure. Although the use of SBUs is quickly emerging as a successful tactic in rationalizing and controlling the formation of various types of extended framework architectures, the use of *s*-block SBUs has not been investigated [17]. In this paper we detail our preliminary investigations in this area, which firmly establish the efficacy of this methodology. We will outline an unprecedented systematic series of one-, two- and three-dimensional polymers constructed from *s*-block metal aggregates. A subset of these networks displays intriguing host-guest chemistry, including the preparation of air-stable, robust, open-framework materials. We have also developed a model for network prediction, which will be an essential element in our future studies this area.

### 2.1 Initial Observations and Proof-of-Concept Preliminary Studies

During the course of our investigations into the structure and function of $\alpha$, $\alpha$'-stabilized carbanions we discovered that the lithium salts of many of these species form unusual polymeric materials [18]. We structurally characterized lithiated $\alpha$-cyanophosphonates of the type $[(RO)_2P(O)CHCNLi.THF]$ (R = Et or $Pr^i$) and found that they form 2D sheet structures, as displayed in **Figure 1**.

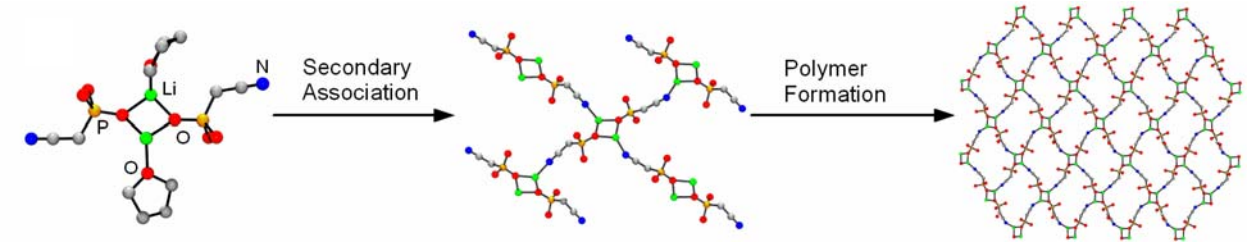



**Figure 1.** Assembly process of lithiated α-cyanophosphonates by linking $Li_2O_2$ dimers.

Intriguingly from the molecular perspective, the sheets can be described as being composed of 'dimeric' $Li_2O_2$ units that are interconnected to four identical 'dimers' via ligation of the nitrile groups. The $Li_2O_2$ motif, where the metals bridge between a pair of phosphoryl units, is entirely consistent with established aggregation patterns for simple lithiated phosphonates [19]. Then, the pendant nitriles act as linear-linkers and bond in an 'interdimer' fashion to another $Li_2O_2$ unit, with the adjacent rings rotated by approximately 90° with respect to one another. Overall, four dimers interconnect through the nitrile bridges to give large $(LiOPCCN)_4$ twenty-four membered rings with small four membered $Li_2O_2$ rings at the corners. This pattern is topologically equivalent to a (4,4) net and can be described as a decorated network where one square-planar vertex has been replaced by a molecular aggregate [20, 21]. The networks obtained are reminiscent of classical structures utilizing square-planar transition metal nodes, such as $Ni^{2+}$, $Zn^{2+}$, $Cu^{2+}$, $Cd^{2+}$, *etc.*, with neutral linear-linker ligands such as 4,4'-bipyridine. In our case the $Li_2O_2$ dimeric aggregates act as square-planar nodes, where the nodal point is located at the center of the ring rather than being an individual atom.

We theorized that it should be possible to form related architectures in a rational and controlled manner using other types of lithiated aggregates. We chose to examine (organo)sulfonylacetonitriles, $RSO_2CH_2CN$, as carbanion sources since simple lithiated sulfonyls are known to form $(SO_2Li)_2$ eight membered ring dimers [22]. Therefore, these complexes should assemble in a similar manner to the cyanophosphonates described above but with larger eight membered rings interconnecting the macromolecular twenty-four membered rings. In this vein we targeted and subsequently successfully prepared and structurally characterized the complex $[MeSO_2CHCNLi.THF]$, [23] and we were pleased to confirm that this complex forms a network with exactly the connectivity that was predicted in advance [24, 25]. **Figure 2** illustrates the formation of this two dimensional sheet that may be described as a basket weave network. Overall, the $(SO_2Li)_2$ eight-membered ring acts as an offset square planar node to give the desired topology.

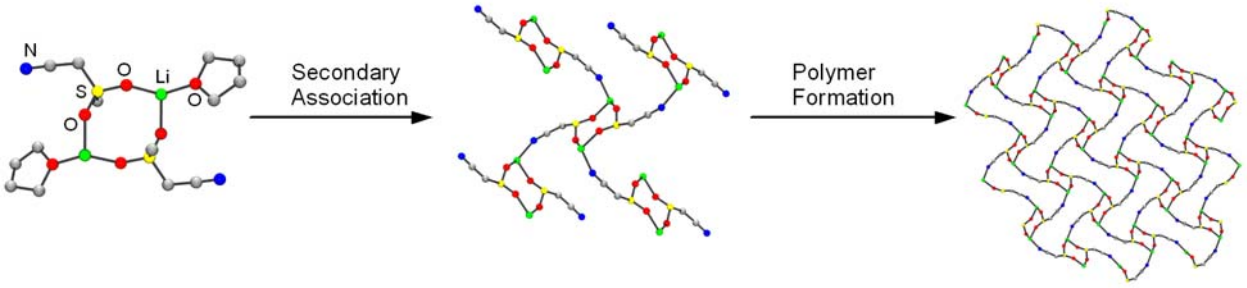



**Figure 2.** Assembly process of lithiated (organo)sulfonylacetonitriles through association of $(SO_2Li)_2$ dimers.

## 3. METHODS

The coordination chemistry involving the early main group elements is an active area of academic study, with the molecular aggregation behavior of numerous classes of metal-ligand combinations now well established [14-16]. With this understanding of molecular aggregation in hand, we aimed to select a series of readily prepared and well-behaved main group aggregates that could then be utilized as SBUs. A highly attractive feature of *s*-block SBUs is the diversity of metal aggregate geometries and compositions available as synthons. Our goal was to utilize the

geometric and chemical information stored in a variety of metal-containing ring and cage compounds to rationally design, and subsequently build, novel two- and three-dimensional network architectures. Specifically, since the location of the metal centers and the anions within these aggregates are known in advance, we could use this information to predict the directions for polymer extension. Suitable aggregates were selected that contain metals with one or more free coordination sites or labile ligands for Lewis base ligation and subsequent polymer growth. In essence the entire aggregate is used as a fixed nodal point (as opposed to the nodes being single spherical metal cations) with their geometries determined by the nature of the rings or cages. This strategy includes utilizing strong 'primary' bonding interactions to assemble molecular aggregates of predictable composition, shape and size, which are then linked via weaker 'secondary' Lewis base-type bonding. The interaggregate association is achieved by the use of neutral polydentate Lewis bases containing at least two donor sites, as opposed to the use of internal donor groups attached to the anionic ligand backbones as described in **Figures 1** and **2**. The use of SBUs in network synthesis is also an appealing approach since they dramatically reduce the number of possible network topologies arising for a given node/linker combination and they may also aid in the preparation of stable porous materials [26].

It is pertinent to recognize that polymer formation is in fact very common for early main group compounds, in particular for *s*-block complexes [27, 28]. Indeed, most preparative chemists have from time to time encountered an insoluble lithiated intermediate during a synthesis. In many instances such solubility problems can be overcome by the use of polar solvents such as THF, DME and TMEDA, which cleave metal-ligand interactions and generally reduce the aggregation state of the metallated species. In turn, the vast majority of work carried out on the structure and bonding of early main group metal complexes has been performed at the molecular level, and the structural elucidation of polymeric materials has mainly been by inference rather than by direct analysis [29]. Although, various types of coordination polymers containing *s*-block metals have previously been characterized, they have usually been prepared either inadvertently or in order to study localized metrical information, rather than deliberately to build specific network architectures [30]. Indeed, while elegant work has been carried out elucidating the subtle factors involved in determining the molecular structures adopted by *s*-block metal complexes, no sustained attempt has been made to rationalize polymer formation in this area.

### 3.1 Common Geometries Available for Alkali Metal Containing SBUs

The use of lithiated SBUs was a reasonable starting point for the study since the understanding of the aggregation behavior for these complexes is by far the most developed within the *s*-block series [14-16]. Our initial phase of the project utilized the abundance of information that is available for lithiated species to choose suitably robust molecular aggregates as building blocks. **Figure 3** displays the core structures of a selection of commonly found aggregation states for *s*-block complexes: ring dimers, ring trimers, tetrameric cubanes, and prismatic hexamers. **Figure 3** also illustrates the relationship between the metal arrangements within the aggregates and their corresponding nodal geometry. The exact arrangement of the metals within each type of aggregate will vary with the identity of its anionic partner. This will lead to a variety of geometric distortions from the 'ideal' nodal configuration. Nevertheless, the basic ring or cage arrangements should remain intact and will therefore control the direction of polymer assembly.

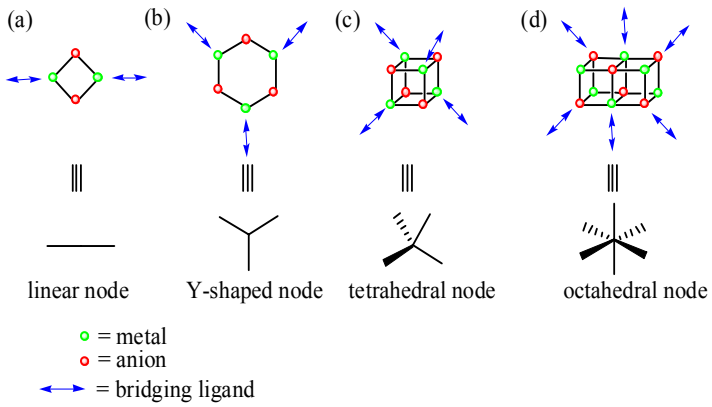



**Figure 3.** Relationship between selected metal aggregates and their related nodal geometries: (a) disolvated ring dimer, (b) trisolvated ring trimer, (c) tetrasolvated tetrameric cubane, and (d) hexasolvated prismatic hexamer.

### 3.2 Lithium Aryloxides as SBUs

It was essential to the success of our approach that the aggregation state of the metallated compound was predictable and remains intact both in solution and in the solid-state while in the presence of the Lewis base linker. An excellent starting point was lithium aryloxides. These complexes contain strong Li-O bonding, and numerous aggregate types are available as SBUs. In addition, they are readily prepared from the parent phenol by direct deprotonation using bases such as organolithium or lithium amide reagents. The aggregate size and coordination environment around the metal center can readily be controlled by the choice of phenol and donor solvent. For example, increasing the steric bulk at the 2,6-positions of an aryloxy ring results in decreasing the aggregation state of the lithiated complex. This is conveniently demonstrated for a series of THF-solvated lithiated aryloxides (ArOLi), shown in **Figure 4**, which form stable ring dimers (Ar = 2,6-$^{t}Bu_2C_6H_3$), ring trimers (Ar = 2,6-$^{i}Pr_2C_6H_3$), tetrameric cubanes (Ar = 2,4,6-$Me_3C_6H_2$), and prismatic hexamers (Ar = $C_6H_5$) [31-33].

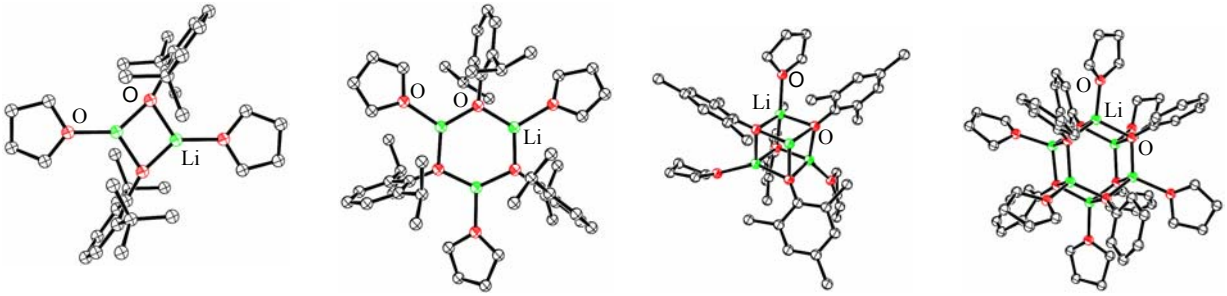



**Figure 4.** Molecular structures of a series of THF-solvated lithium aryloxides, $[ArOLi.THF]_n$.

Each of the metal atoms within these aggregates has one coordination site occupied by a monodentate THF donor molecule. This makes these anionic aryloxide ligands excellent targets as SBUs since replacement of the terminal Lewis bases by didentate linear-linker molecules will result in interconnecting aggregates. There are of course numerous possibilities for both the connectivity and topology of the supramolecular assemblies that may be formed via the interaction of the four aggregate types shown above with a linear-linker such as dioxane. However, it is becoming clear that a relatively few, high-symmetry structures dominate network topologies [26]. The most likely assemblies to be produced in each case are one-dimensional chains from disolvated dimers; hexagonal (6,3) sheets from ring trimers; three-dimensional diamondoid networks using tetrameric cubanes; and cubic lattices using prismatic hexamers.

### 3.3 Choice of Neutral Linker Ligand

Polydentate Lewis bases are required in order to connect neighboring aggregates. The strategy used to achieve this goal was to employ neutral linear-linker ligands containing relatively hard Lewis base donors [34]. A selection of potentially useful linkers for the *s*-block systems are shown in **Figure 5**.

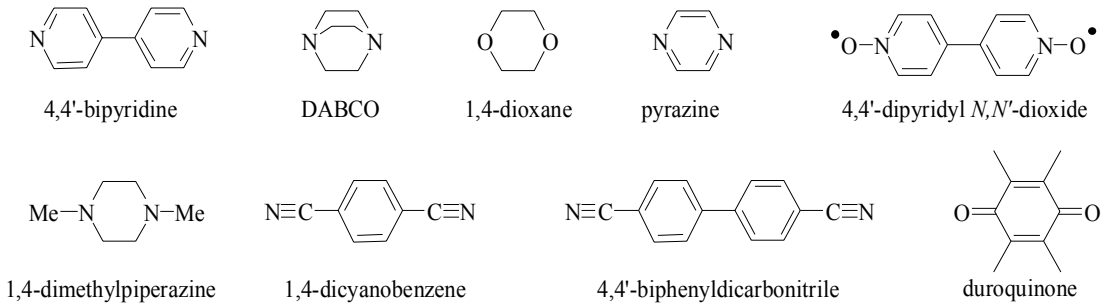



**Figure 5.** Neutral linear-linker molecules suitable for coordination to hard *s*-block metals.

Generally, divergent ligands have been used in lithiated systems simply to enhance the crystallization properties of a metallated molecular substrate in order to study localized bonding or metrical information, rather than to

deliberately build specific network structures [30, 35, 36]. However, in most instances no recognizable lithiated aggregate is present, rather isolated ions with no inherent directional control are linked together. Furthermore, reports on the use of other linear-linking didentate Lewis bases for the *s*-block metals are surprisingly scarce. This deficiency is remarkable considering the enormous importance of such ligands in the assembly of transition metal-containing networks [37]. Nevertheless, ligands of this type should be ideal candidates for coordination to *s*-block metals, in particular lithium, due to their excellent properties as hard Lewis bases. The monodentate base pyridine is a commonly used donor solvent for lithiated complexes with dozens of structurally characterized examples in the Cambridge Structural Database [38], and furthermore several crystal structures of lithium species have been characterized containing the chelating ligand 2,2'-bipyridine [39]. In the present study the linker of choice is 1,4-dioxane as this ligand proved to be convenient as it not only acts as a divergent linker but also as solvent media.

## 4. RESULTS AND DISCUSSION

### 4.1 Use of Tetrameric $Li_4O_4$ Cubanes as SBUs Linked by Dioxane

This early success in rationally designing complex network architectures from very simple constituents spurred us to expand the scope of our studies to include the possibility of linking together neutral, pre-assembled lithiated SBUs using *external* neutral divergent Lewis bases. This is appealing since it immediately opens up a wide variety of well-studied solvated aggregate types for investigation. We targeted the commonly encountered $Li_4O_4$ tetrameric cubanes formed by lithium aryloxides (ArOLi) for use as SBUs [31]. These complexes are excellent SBU candidates since they contain strong Li-O bonding, the metals are held in an approximately tetrahedral arrangement with one coordination site available for ligation, and there is a wide range of substituted phenols available for systematic studies. The didentate donor dioxane was selected as the divergent external linking Lewis base due to its capacity to act as solvent media, its rigidity and its inability to chelate a lithium center.

At the outset, an exploratory computational study was conducted to determine the feasibility of dioxane acting as a bridging ligand between a pair of $Li_4O_4$ cubanes, since this structural pattern had not previously been characterized (**Figure 6**). Semi-empirical PM3 calculations were the first line of analysis due to their computational efficiency [40, 41]. Geometry optimization calculations were performed on the full molecules [(PhOLi.dioxane)$_4$] and [{(PhOLi)$_4$}$_2$.(dioxane)$_7$] rather than using simplified model complexes to ensure an accurate portrayal of subtle steric and electronic effects. These calculations indicated that no notable energy penalty (<0.2 kcal/mol) is incurred on bridging a pair of tetrasolvated cubanes by the donor. This key result was confirmed by further geometry optimization at the *ab initio* HF/6-31G* level of theory (<0.01 kcal/mol) and also by high level single-point density functional theory calculations (B3LYP/6-311G**, <0.04 kcal/mol) [41-44]. These calculations firmly established the viability of polymer formation. In addition, this study demonstrates the ability of lower level calculations to be used as an efficient method for screening suitable aggregate/linker combinations.

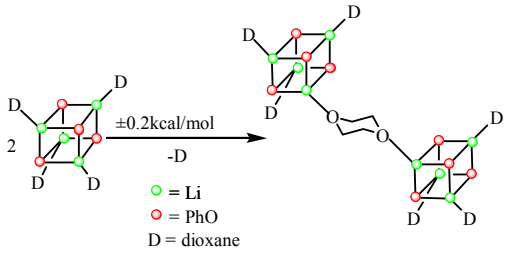



**Figure 6.** Calculated energetics of dioxane bridging a pair of $Li_4O_4$ units.

Appropriate lithium aryloxides were then chosen based on their predilection to form tetrasolvated $Li_4O_4$ cubanes in the presence of monodentate Lewis bases [23]. The required tetrameric aggregation was confirmed by preparing and structurally characterizing (by X-ray diffraction, XRD) the molecular analogues using monodentate donors (THF or pyridine), and also by solution NMR studies of the complexes in dioxane solvent media using the Jackman method of correlating the $^{13}C$ NMR shift positions for the *para*-carbons with aggregation state [25, 45]. Specific structural variants of these ligands were then selected as our understanding of these systems evolved. **Figure 7** highlights the structural characterization, by single-crystal XRD, of the three representative complexes [{(ROLi)$_4$.(dioxane)$_x$}$_\infty$], (a) R=Ph, x=3; (b) R=4-Et-$C_6H_4$, x=2.5; and (c) R=1-naphth, x=2, which typify our findings [46].

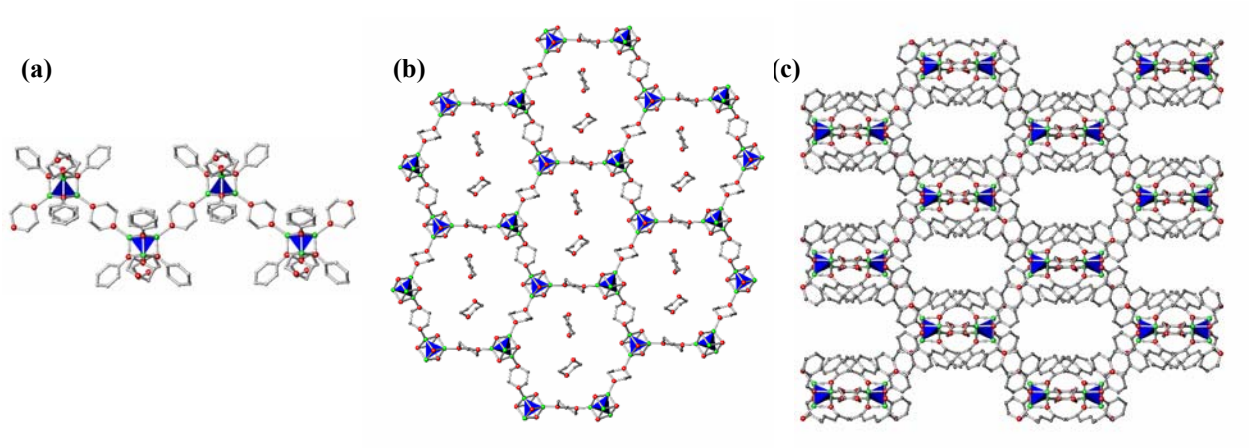



**Figure 7.** Sections of the polymeric structures of the: **(a)** 1D zig-zag chain, **(b)** framework atoms of the 2D hexagonal network with two guest solvent molecules per macrocycle, and **(c)** 3D diamondoid network. Vertices of the blue tetrahedra represent the four lithium centers within each cubane.

The first key point is that each structure is composed of linked $Li_4O_4$ tetrameric cubane units, proving that the desired molecular SBU remains intact on polymerization. Next, the three polymeric types found, 1D zig-zag chains, 2D hexagonal nets, and 3D diamondoid networks are very closely related, simply differing by the number of points of extension from the tetrahedral SBU (two, three and four leading to V-shaped, trigonal and tetrahedral nodes respectively). Moreover we have now fully characterized an extended series of $Li_4O_4$ substituted aryloxide complexes and found that they each form one of these three structural types. Pleasingly, this demonstrates that the SBU approach is indeed successful in limiting the number of network topologies for a given node/linker combination.

The complexes in **Figure 7** illustrate the dramatic effect of relatively small changes in sterics on the macromolecular structure adopted. We deduced that the type of polymer formed can be rationalized in terms of balancing entropic factors with the efficient filling of space. Although our calculations indicate that bridging by dioxane is thermoneutral, there will be a substantial increase in entropy associated with polymer formation due to the gain in translation freedom of the liberated solvent molecules. Next, there will clearly be an optimum length for the rigid anionic ligands of the SBUs in order to efficiently fill the hexameric macrocycles of the 2D sheets. If the ligands are too short this will result in energetically unfavorable large voids at the center of the rings. Conversely, if the ligands are too long the 2D arrangement will be destabilized by transannular interactions [47]. This analysis is consistent with the 1D chain structure observed for the small $PhO^-$ anions. Extension of the ligand at the *para*-position partially circumvents this problem and the longer 4-Et-$C_6H_4O^-$ anions, in combination with the guest dioxanes, fill the hexameric macrocycles with reasonable efficiency. However, examination of the extended crystal-packing diagram of this complex shows interdigitation between neighboring 2D layers, suggesting that significant space still remains within the sheet. In order for a diamondoid structure to be formed it is necessary to occupy 3D adamantanoid cavities. This proves possible for 1-naphthOLi by lateral extension to the aromatic ring, leading to excellent localized space filling in the 'corners' of the adamantanoid cavities by interlocking of the naphthyl units [47]. Space filling by interpenetration is precluded due to the $(1\text{-naphthOLi})_4$ cubanes being ~14.2 Å in diameter whereas the adamantanoid cavities have dimensions of only 7.3 x 7.6 x 16.6 Å. The open structure found for this complex supports the conjecture that the use of SBUs promotes the formation of open-framework materials. The viability and scope of our rationalization was tested through targeting the structural characterization of specific substituted aryloxides. This proved successful, such that when the *para*-substituent is small (H, Me, F or Cl) 1D chains are produced, minimal extension at this position (Et, OMe, $^i$Pr) leads to the formation of 2D hexagonal sheets, whereas lateral substitution of the aromatic (Ar = 1-naphth or 2,4,6-$Me_3C_6H_2$) gives 3D diamondoid networks. In all cases the homogeneity of the samples was confirmed by powder XRD of the bulk solids.

The potential for porosity is an exciting feature of the diamondoid structures outlined above and our preliminary investigations indicate that the guest molecules within $[\{(1\text{-naphthOLi})_4.(\text{dioxane})_2\}_\infty]$ (**Figure 7c**) can be removed under reduced pressure without disruption to the framework. Solvent loss was monitored by $^1$H NMR spectroscopy through complete dissolution of samples in $d_6$-DMSO, followed by integration of the anion and dioxane signals.

Initially, three enclatherated solvent molecules per $Li_4O_4$ unit are present but after 24 hours of evacuation ($1.5 \times 10^{-3}$ Torr and 30°C) the non-framework dioxane is entirely removed. Calculations indicate that 34.8% of the total volume within this structure is potential solvent space.[77] Comparison of the calculated, as prepared, and evacuated powder XRD patterns show only minor variations, demonstrating that the integrity of the framework is maintained on solvent loss. In addition, the macroscopic morphology of the crystals is retained after evacuation (perfect octahedra) allowing determination of the unit cell parameters by single-crystal XRD, and pleasingly they match those of the original material (a full structure determination has not yet been completed due to the weak diffraction of the crystals). Furthermore, these crystals were found to be completely air-stable over several weeks. Unit cell checks confirmed that the crystallinity is retained and IR spectroscopy showed no detectable decomposition to lithium hydroxide.

### 4.2 Use of Hexameric $Na_6O_6$ Prismatic Cages as SBUs Linked by Dioxane

Many complexes of the heavier Group 1 metals are appealing as SBUs. Of the heavier alkali metals, the alkoxides and aryloxides of both sodium and potassium have been reasonably well studied [48, 49]. A potential drawback of moving from lithium to the heavier Group 1 metals is a significant weakening of the metal-ligand bond strength, which may in turn lead to increased aggregate lability and a decrease in the predictability of the nature of the SBU. To test the feasibility of using these metal aggregates we targeted solvated $Na_6O_6$ hexameric aggregates as potential sodium based SBUs since this structural type has been characterized in the solid state for molecular derivatives, *e.g.* $[(PhONa.THF)_6]$ [50]. We theorized that the triple stack of dimers (face-shared cubane) aggregates should act as octahedral SBUs and lead to the formation of 3D cubic networks. Building on our work on the lithiated systems, we utilized 4-substituted aryloxides as convenient probes to test the effect of sterics on the supramolecular arrangements obtained. The set of *para*-halide substituted derivatives are particularly useful in illustrating the effect of sterics on network structure. The complexes $[(4\text{-R-}C_6H_4ONa)_6.(dioxane)_x]_\infty$, where (a) R = I, x = 5 form 1D linear chains, (b) R = Br, x = 4 from (4,4) 2D square nets, and (c) R = F and x = 3 form 3D cubic lattices, **Figure 8** [51]. This remarkable set of complexes again demonstrates that the desired hexameric $Na_6O_6$ SBU is robust and remains intact within each network. Also, the type of network can again be rationalized in terms of the size of the substituent at the *para* position, with only the fluorine group being small enough to fit into the available volume within the cubic cavities. Moreover, we have again confirmed that network formation using the $Na_6O_6$ SBUs is limited to these three structural types, *i.e.* structural characterization of the dioxane solvated *para*-aryloxides R-$C_6H_4ONa$: R = Cl, I or Ph, give 1D linear chains, R = Br, H, Me, $^i$Pr, $^t$Bu or $^t$Pn give 2D (4,4) nets and R = F or Et give 3D cubic networks. These results show that even the heavier *s*-block elements may be used as robust SBUs.

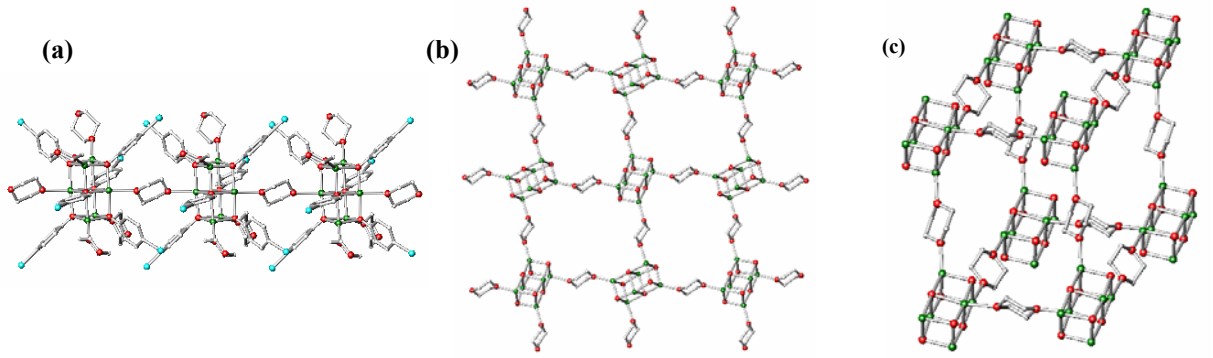



**Figure 8.** Sections of the polymeric structures of the: (a) 1D linear chain, **(b)** 2D square net (framework atoms only), and **(c)** 3D cubic lattice (framework atoms showing a single cubic unit), formed from triple-stack $Na_6O_6$ SBUs bridged by dioxane.

## 5. CONCLUSIONS

To summarize, we have demonstrated that both lithiated and sodiated pre-assembled molecular aggregates may be used to construct network assemblies. The topologies of the resulting frameworks are directed by the geometry of the metals within the molecular SBUs and is also dependant on the ability of organic components to fill space effectively. These initial studies suggest that there is a rich chemistry surrounding the use of early main group metals in the supramolecular synthesis of extended frameworks.

## 6. ACKNOWLEDGEMENTS

We gratefully acknowledge the Petroleum Research Fund (41716-AC3) and the University of Notre Dame for support. We also thank the National Science Foundation for instrumentation support (CHE-0443233).

## 7. REFERENCES

[1] L. Brammer, *Chem. Soc. Rev.* **33** (2004) 476.
[2] D. Braga, *Chem. Commun.* (2003) 2751.
[3] S. L. James, *Chem. Soc. Rev.* **32** (2003) 276.
[4] C. Janiak, *Dalton Trans.* (2003) 2781.
[5] J. K. M. Sanders, *Chem. Eur. J.* **4** (1998) 1378.
[6] G. B. Gardner, D. Venkataraman, J. S. Moore and S. Lee, *Nature* **374** (1995) 792.
[7] S. J. Lee and W. Lin, *J. Am. Chem. Soc.* **124** (2002) 4554.
[8] H. O. Stumpf, L. Ouahab, Y. Pei, D. Grandjean and O. Kahn, *Science* **261 (**1993) 447.
[9] J. W. Steed and J. L. Atwood, *Supramolecular Chemistry*, (Wiley, New York, 2000).
[10] G. R. Desiraju, *Crystal Engineering. The Design of Organic Solids*, (Elsevier, Amsterdam, 1989).
[11] I. Haiduc and F. T. Edelmann, *Supramolecular Organometallic Chemistry*, (Wiley-VCH, New York, 1999).
[12] V. Chandrasekhar, S. Nagendran, S. Bansal, A. W. Cordes and A. Vij, *Organometallics* **21** (2002) 3297.
[13] O. M. Yaghi, M. O'Keeffe, N. W. Ockwig, H. K. Chae, M. Eddaoudi and J. Kim, *Nature* **423** (2003) 705.
[14] T. Stey and D. Stalke, D. in: *The Chemistry of Organolithium Compounds*, eds. Z. Rappoport and S. Patai (Wiley, New York, 2004), p 47.
[15] A. M. Sapse and P. v. R. Schleyer, *Lithium Chemistry, A Theoretical and Experimental Overview*, (Wiley, New York, 1995).
[16] E. Weiss, *Angew Chem. Int. Ed. Engl.* **32** (1993) 1501.
[17] An excellent special addition of the Journal of Solid-State Chemistry is available: O. M. Yaghi, M. O'Keefe, eds. *J. Solid State Chem.* **152** (2000) 1-321.
[18] K. W. Henderson, A. R. Kennedy, A. E. McKeown and D. S. Strachan, *J. Chem. Soc., Dalton Trans.* (2000) 4348.
[19] S. E. Denmark, K. A. Swiss and S. R. Wilson, *Angew. Chem. Int. Ed. Eng.* **35** (1996) 2515.
[20] A. F. Wells, *Three-Dimensional Nets and Polyhedra*, (Wiley-Interscience, New York, 1977).
[21] J. Kim, B. Chen, T. M. Reineke, H. Li, M. Eddaoudi, D. B. Moler, M. O'Keeffe and O. M. Yaghi, *J. Am. Chem. Soc*. **123 (**2001) 8239.
[22] G. Boche, *Angew. Chem. Int. Ed. Engl.* **28** (1989) 277.
[23] K. W. Henderson, A. R. Kennedy, L. Macdonald and D. J. MacDougall, *Inorg. Chem.* **42** (2003) 2839.
[24] K. W. Henderson, A. R. Kennedy and D. J. MacDougall, *Inorg. Chem.* **42** (2003) 2736.
[25] K. W. Henderson, A. R. Kennedy and D. J. MacDougall, *Phosphorous, Sulfur and Silicon* **179** (2004) 795.
[26] N. W. Ockwig, O. Delgado-Friedrichs, M. O'Keeffe and O. M. Yaghi, *Acc. Chem. Res.* **38** (2005) 176.
[27] D. Braga, F. Grepioni and G. R. Desiraju, *Chem. Rev.* **98** (1998) 1375.
[28] B. Walfort, L. Lameyer, W. Weiss, R. Herbst-Irmer, R. Bertermann, J. Rocha and D. Stalke, *Chem. Eur. J.* **7** (2001) 1417.
[29] R. E. Dinnebier, U. Behrens and O. Falk, *J. Am. Chem. Soc.* **120** (1998) 1430.
[30] For a specific example of 1,4-dioxane bridging in a lithiated coordination polymer see: O. Kühl, J. Sieler, G. Baum and E. M. Hey-Hawkins, *Z. Anorg. Allg. Chem.* **626** (2000) 605.
[31] T. J. Boyle, D. M. Pedrotty, T. M. Alam, S. C. Vick and M. A. Rodriguez, *Inorg. Chem*. **39** (2000) 5133.
[32] K. Thiele, H. Gorls and W. Seidel, *Z. Anorg. Allg. Chem*. **624** (1998) 1391.
[33] L. M. Jackman, D. Cizmeciyan, P. G. Williard and M. A. Nichols, *J. Am. Chem. Soc.* **115** (1993) 6262.
[34] P. K. Chattaraj, H. Lee and R. G. Parr, *J. Am. Chem. Soc.* **113** (1991) 1855.
[35] F. T. Deelmann, F. Pauer, M. Wedler and D. Stalke, *Inorg. Chem.* **31** (1992) 4143.
[36] M. Motevalli, D. Shah and A. C. Sullivan, *J. Chem. Soc, Dalton Trans*. (1993) 2849.
[37] R. Robson, in: *Comprehensive Supramolecular Chemistry*, eds. J. L. Atwood, J. E. D. Davies, D. D. MacNicol, F. Vögtle and K. S. Suslick, (Elsevier, Oxford, 1996).
[38] F. H. Allen, *Acta Crystallogr.* **B58 (**2002) 380.
[39] For example see: B. W. Skelton, C. R. Whitaker and A. H. White, *Aust. J. Chem.* **43** (1990) 755.

[40] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, C. G. Zakrzewski, J. A. Montgomery Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman [illegible]7[illegible]cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, [illegible] Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, J. L. Andres, M. Head-Gordon, E. S. Replogle and J. A. Pople, *Gaussian 98*, revision A.6; Gaussian, Inc.: Pittsburgh, PA, 1998.
[41] W. J. Hehre, R. Ditchfield and J. A. Pople, *J. Chem. Phys.* **56** (1972) 2257.
[42] P. C. Hariharan and J. A. Pople, *Theo. Chem. Acta* **28** (1973) 213.
[43] A. D. Becke, *J. Chem. Phys.* **98** (1993) 5648.
[44] A. D. McLean and G. S. Chandler, *J. Chem. Phys.* **72** (1980) 5639.
[45] L. M. Jackman and B. D. Smith, *J. Am. Chem. Soc.* **110** (1988) 3829.
[46] D. J. MacDougall, J. J. Morris, B. C. Noll and K. W. Henderson, *Chem. Commun.* (2004) 456.
[47] S. Mecozzi and J. Rebek Jr., *Chem. Eur. J.* **4** (1998) 1016.
[48] R. E. Dinnebier, M. Pink, J. Sieler and P. W. Stephens, *Inorg. Chem.* **36** (1997) 3398.
[49] T. J. Boyle, N. L. Andrews, M. A. Rodriguez, C. Campana and T. Yiu, *Inorg. Chem.* **42** (2003) 5357.
[50] M. Kunert, E. Dinjus, M. Nauck and J. Sieler, *Chem. Ber.* **130** (1997) 1461.
[51] D. J. MacDougall, B. C. Noll and K. W. Henderson, *Inorg. Chem.* **44** (2005) 1181.

# MODELING STUDIES OF ANTI-AMYLOID ANTIBODIES BOUND TO Aβ

Anna S. Gardberg, Nathan Tanner, Chris Dealwis*
*University of Tennessee at Knoxville, Department of Biochemistry and Molecular Biology, M407 Walters Life Sciences*
*Knoxville, Tennessee 37996-0830*

Israel Huff, Kimberly Salone, Ronald Wetzel
*University of Tennessee at Knoxville, Graduate School of Medicine, Knoxville, Tennessee, USA*
*Corresponding author: cdealwis@utk.edu
Keywords: A-beta, beta-amyloid, monoclonal antibodies, Alzheimer's, neurodegenerative disorders Autodock, protein-protein interaction
Grant sponsors: AG, NT, IH, RW, and CD are supported by NIH grant 5R01NS046356-03. KS was supported by a grant from the American Health Assistance Foundation.

## 1. ABSTRACT

WO1 and WO2 are conformation-specific monoclonal IgMs that bind the fibril state of the amyloid Aβ peptide (1-40), as well as amyloid fibrils of other disease-related proteins. Significantly, these antibodies (Abs) do not bind the soluble, monomeric state of Aβ (1-40) or the precursor form of other amyloids. The Abs have been sequenced and compared with sequences in the Kabat database; though some unusual charged residues were noted in WO1 and WO2, none are at the CDR. Three-dimensional models of the Fv fragments of WO1 and WO2 were generated with Web Antibody Modeling. A trigonal prism and a two-strand structural model of the Aβ amyloid core were compared by docking each with the Fv models of WO1 and WO2. The results predict binding of WO1 and WO2 to the Aβ trigonal prism at the protofilament face consisting of residues 23-27 and of WO2 to the Aβ two-strand at the end of the elongating fibril. No positive prediction for WO1 and the Aβ two-strand was made. Binding experiments with WO1 were determined to be sensitive to salt and pH conditions, suggesting the importance of electrostatic interactions for binding; this result is consistent with the proposed docking of WO1 to the Aβ trigonal prism.

## 2. INTRODUCTION

The aggregation of normally soluble proteins into insoluble, unbranched fibrils is the underlying pathology of a family of diseases known as the amyloidoses.[1] The hallmark event in amyloidogenesis is a change in the secondary and/or tertiary structure of a normal, soluble protein, rendering it prone to self-assembly into highly ordered para-crystalline arrays: fibrils. More than 20 proteins have been clinically identified as precursors of amyloid fibrils *in vivo*. These include the amyloid precursor protein (APP), Islet amyloid polypeptide (IAPP), α-synuclein, transthyretin (TTR), immunoglobulin light chain (LC), polyglutamine-repeats, and prion proteins, that are associated with diseases such as Alzheimer's,[2-4] type II diabetes,[5,6] Parkinson's disease,[7-9] familial polyneuropathy,[10] light chain associated (AL) amyloidosis,[11-13] Huntington's disease,[4,14] and the spongiform encephalopathies.[15] By understanding the three-dimensional structure of such fibrils, we might design therapeutic agents to target them. It is notoriously difficult to extract structural information directly from amyloid fibrils, which are insoluble and non-crystalline, so a great deal of study has gone into fiber diffraction studies,[16-18] mutation studies,[19] microscopy,[17,20,21] proteolysis,[22,23] ESR,[24,25] NMR,[24,26] SANS,[27] and deuterium exchange with mass spectroscopy[28-31] in an ongoing attempt to elucidate the molecular structure of the Aβ fibrils associated with Alzheimer's disease.

There is no consensus model for Aβ structure, but most models incorporate cross-beta secondary structure. The cross-beta structure (in which the fibril axis is perpendicular to the chain direction) of Aβ models is derived from the 4.75 Å meridional reflection in fiber diffraction studies.[18] A stacked, parallel beta-sheet arrangement is suggested by Burkoth, Benzinger, and others,[24,27,32] while a pair of concentric cylinders was proposed by Perutz, *et al*.[33] An anti-parallel or stacked-hairpin beta-sheet arrangement is suggested by Callaway and others.[34-37]

One recent model of the Aβ protofilament, proposed by Guo, *et al,*[38] features a trigonal prism of stacked, parallel Aβ(15-36) polypeptides; the model is shown in **Figure 1a**. This trigonal prism protofilament model is consistent with threading analysis[38] and proline-scanning mutagenesis,[19] and the proposed fibril model (of 6 protofilaments) appears to be consistent with electron microscopy. Recent data from hydrogen exchange NMR[39] and cysteine scanning mutagenesis[40] have called portions of this construction into question.

Another proposed configuration of the Aβ protofilament consists of two stacks of two parallel strands twisting around one another along the fibril axis; this model, shown in **Figure 1b**, is consistent with data that was available in 2002, especially that from solid-state NMR studies.[26] It is also in line with the recent crystal structure of the amyloid form of a seven-residue fragment from yeast protein Sup35[41], which supports a double-beta-sheet structure. Both the two-strand and trigonal prism Aβ protofilament constructs are consistent with specific bodies of experimental data, but are not completely determined by experimental data. Both models continue to evolve with the acquisition of new data.

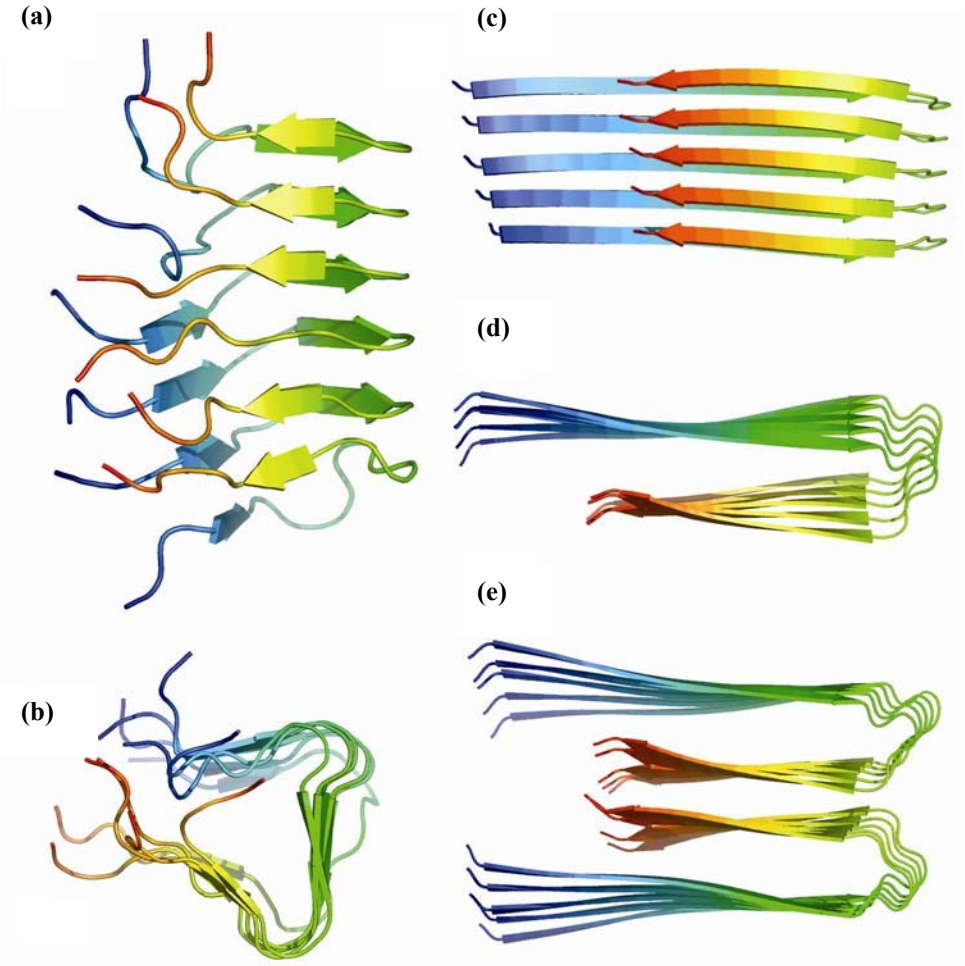



**Figure 1. (a)** Three-strand protofilament model in side view. **(b)** Three-strand, top view. **(c)** One half of proposed two-strand Aβ, side view. **(d)** Top view. **(e)** Two-fold symmetry about the *x*-axis has been used to generate one proposed configuration for the full protofilament from the two-strand model.

The monoclonal antibodies WO1 and WO2 (κ-light chain, murine, IgM) recognize a common conformational epitope shared by several different types of amyloid fibrils, with little dependence on amino acid sequence; the two bind the same antigen with similar affinities for the Aβ fibril.[42] This result opens the door to targeting whole classes of amyloid fibrils for detection and therapy.[43] Work to crystallize these antibodies (or their fragments) is ongoing. While IgM's are notoriously difficult to model,[44] the success of antibody modeling algorithms at predicting crystal structures of antibody variable regions[45] encouraged the use of computational models of WO1 and WO2 Fvs for docking simulations. Since WO1 and WO2 bind to Aβ amyloid, we hypothesized that a model of

the Aβ protofilament would, if correct, dock *in silico* to a model of the variable regions of WO1 and WO2. The configuration of the predicted complexes of the Fvs of antibodies WO1 and WO2 to two different models of Aβ protofilament is given. Experimental salt-dependence data is consistent with the prediction of WO1-Aβ binding for the trigonal prism model.

## 3. METHODS

### 3.1 Sequence Analysis

The nucleotide sequences of WO1 and WO2 were obtained by cloning and confirmed by multiple cloning using high fidelity polymerases, and further confirmed by extended N-terminal amino acid sequencing from analysis of the protein. Initial alignments were performed using Molecular Operating Environment (MOE™[46]), with Ab sequences obtained from the Protein Data Bank[47,48]. An extensive alignment was done using the Kabat sequence database testing program[49]. BLASTP analysis[50] was run on the WO1 and WO2 variable light chains using the Non-Redundant Protein Database at the San Diego Supercomputer Center. Canonical classes are based on Chothia nomenclature[51] and numbering scheme is that of Kabat[52]. The WO1 and WO2 Fv residue sequences were initially compared to one another. Then they were compared with sequences from the Kabat database to search for anomalies, unique features, or homologies to other antibodies. A sequence alignment of the CDRs is shown in Figure 2.

### 3.2 Model Building

Three-dimensional structures of WO1 and WO2 variable regions (Fv) were generated using the Web Antibody Modeling algorithm, WAM.[45] WAM is an improvement on the AbM program [53-56] with greater capability in modeling the highly variable H3 loop through a combination of knowledge-based and *ab initio* methods[45] . The WAM algorithm has a record of producing models that are 1.0-2.8 Å RMSD from observed structures for the heavy chain CDR3, and better than that for the canonical loops (typically 1.0-2.5 Å RMSD).[45]

### 3.3 Docking

#### 3.3.1 Preparation

A PDB file of the trigonal prism Aβ 1-40 hexameric protofilament model after molecular dynamics simulations ("Aβ trigonal prism", which models residues Gln15 through Val36)[38] was kindly provided by Juntao Guo and Ying Xu. A PDB file of the two-strand pentameric Aβ 1-40 protofilament model after energy minimization ("Aβ two-strand", which models residues 9 through 40)[26] was kindly provided by Robert Tycko.

The program Autodock3[57] predicts the interaction of ligands with macromolecular targets. As Autodock3 can handle no more than 2048 atoms, we reduced the two-strand protofilament to a tetramer and did not generate its symmetry pair. Control calculations performed with a protein-peptide complex, a protein-ligand complex, and a lysozyme-anti-lysozyme complex of known structure (PDB codes 1SMR, 1DQJ and 1DYI) showed that the charges assigned from forcefield calculations performed by AutoDockTools yielded less accurate results than those assigned by AMBER forcefield calculations[58,59] performed by InsightII.[60] We assigned charges using the following protocol. The atom names were standardized to Refmac5 conventions. Using Refmac5[61,62], hydrogen atoms were added in “riding” positions computed from the carbon, nitrogen, oxygen, and sulfur atomic positions. After capping the termini of each chain with (neutral) carboxylic and amine end groups, the partial charges for each atom (including polar hydrogen atoms) in each model were calculated with the InsightII program using the AMBER force field. Calculating the summed formal and partial charges for each model verified that the model is electrically neutral. A Sybyl-style file was then exported from InsightII.

#### 3.3.2 Computation

Autodock-style PDBQ files were prepared from the Sybyl-style files (described in the previous section) with *mol2topdbq* and *mol2topdbqs*, awk-based utilities packaged with Autodock3.[57] For purposes of solvation calculation, we designated the Fv models the “proteins” and the Aβ protofilament models the “ligands”. Despite the thermodynamic importance of side-chain motion,[63] neither the Fvs nor the Aβ models were permitted any torsional freedom; the Autodock package does not permit the side chains of the protein to move, and the Autotors utility for designating rotatable bonds in the ligand allows a maximum of 32, not nearly sufficient for a the 2395-2040-atom Aβ models. Grid parameter files were generated *via mkgpf3* (a script which comes packaged with

Autodock3) and adjusted to compensate for the large volume required for the interaction of the Fab's complementarity determining region (CDR) with the protofilament models. The grid maps were calculated using AutoGrid. The gridded area dimensions were 66 Å x 66 Å x 66 Å, with grid spacing increased to 0.55 Å. The final gridded regions encompassed the CDR of the Fv (as well as all of the heavy chain and most of the light chain of the model for WO1, and all save the C-termini for WO2) and the starting position of the Aβ model, with room for rotation and translation. Grid generation was performed with Autogrid3.
Docking was performed using the Lamarckian genetic algorithm (LGA), and the pseudo-Solis and Wets methods were applied for the local search. Each docking experiment was performed 50 times. Default parameters were applied for the docking experiments, with the exception of the initial rotation step size, which was set to 180. The results of the docking experiments were evaluated by calculating the positional root-mean-square (rms) deviation of the corresponding atoms of each conformation. Docking parameter files were prepared with the *mkdpf3* setup utility and adjusted by hand.

### 3.3.3 Contact Analysis

Contacts between the Fv models and the protofilament models in the predicted docking complexes were analyzed with the program CONTACT.[62] Docking clusters and configurations were examined Pymol,[64] with residue charge surfaces visualized. Electrostatic surface plots for publication were generated by the DelPhi module of InsightII.[60] Angles between solutions were computed by calculating the average direction vector along the protofilament axis for each solution and taking the inverse cosine of the dot-product of direction vectors.

### 3.4 Electrostatic Analysis

Electrostatic surfaces were calculated using DelPhi, a part of the InsightII graphical software package. The following parameters were employed: solvent dielectric 80, radius 1.4, ion radious 2.0, grid solvent extent 15, 65 points. Models were displayed and manipulated with InsightII and all computational work done on a Silicon Graphics computer system.

### 3.5 Salt Effects on Binding WO1

Binding of the WO1 antibody to Aβ fibrils was tested under 13 different conditions of cation, anion, salt concentration, and pH. 1xPBS buffer was used as a control. Experiments 1-5 comprised the salt-effects experiments; they consisted of: 5 mM HEPES at pH 7.5 with: 1) no salt, 2) 150 mM NaCl, 3) 600 mM NaCl, 4) 150 mM KCl, and 5) 75 mM $Na_2HPO_4$. In experiments 6-13, the effects of pH were examined; they consisted of:150 mM NaCl with 6) citric acid pH 3.0, 7) citric acid pH 5.8, 8) PIPES at pH 5.8, 9) PIPES at pH 7.4, 10) Bicine at pH 7.4, 11) Bicine at pH 8.5, 12) CHES at pH 8.5, and 13) CHES at pH 10.0.

No ionic interference was expected from HEPES, which has low ionic strength. $Na_2HPO_4$ has roughly double the ionic strength of the other two salts so it was used at half the concentration for a balanced comparison. The salt-effect experiments were buffered to pH 7.5, near physiological pH conditions. The pH effect experiments contained 150 mM NaCl to provide approximate physiological ionic strength conditions. The Aβ fibrils were prepared as described in *Kheterpal, et al, 2000*.[31]

## 4. RESULTS

### 4.1 Sequence Analysis

### 4.1.1 Comparison of WO1 to WO

The light chain variable regions share 66% identity (80% similarity) and the heavy chain variable regions share 82% identity (85% similarity). Overall charge composition of the Fv fragments is very similar, with isoelectric points (pIs) of the variable regions calculated to be: WO1 $V_H$ 7.3, $V_L$ 8.7, WO2 $V_H$ 7.2, and $V_L$ 8.7.

While both sequences are rich in hydrogen bonding residues, WO1 $V_L$ CDR has a greater proportion of charged Asp, Glu, Arg, and Lys residues (19% vs. 11%), while the WO2 $V_L$ CDR is richer in uncharged, polar hydroxyl residues (42% for WO1 $V_L$, 64% for WO2 $V_L$). The sequences of the CDR's are compared in **Figure 2**. The WO1 and WO2 heavy chain variable regions are more similar in overall sequence, and the number of charged residues in the $V_H$ is identical. Still, the two charged residues of WO1's $V_H$ are at its CDR, and those of WO2's $V_H$ are not at its CDR. WO1's light chain CDR contains charged residues at positions occupied by hydroxyl residues in WO2 such as Arg24, Lys52, and Asp56 in WO1 which are replaced by Thr, Ser, and Ser in WO2. Similarly, the unique charged

residue Arg93 in WO2 replaces a serine residue in WO1. WO1's greater proportion of charged residues and WO2's greater proportion of uncharged polar residues suggest that WO1 and WO2 bind amyloid differently from one another.

a)

| VL CDR1 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 31a | 32 | 33 | 34 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| WO1 | R | A | S | G | N | I | H | N | - | ***Y*** | L | A |
| WO2 | T | A | S | S | S | V | S | S | S | ***Y*** | L | H |
| α-Lyso | R | A | S | Q | N | I | S,H | N | N | - | L | H,A |
| Germ. | R,K | S,A | S | Q | S,N | I | nh* | nh | S | nh | L | nh |
| Kappa | R,K | A,T | S | Q | nh | V,I | nh | S,T | - | nh | L | nh |

| CDR2,3 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| WO1 | N | A | K | T | L | ***A*** | D | Q | H | F | W | S | T | P | ***Y*** | T |
| WO2 | S | T | S | N | L | ***A*** | S | L | Q | Y | H | R | S | P | ***Y*** | T |
| α-Lyso | Y | T | T,S | Q,T | S,L | nh | D,S | Q | Q | F,Y | nh | S | - | P | R,L | T |
| Germ | nh | A,V | S,A | nh | R,L | D,E | S | nh | Q | nh | S | nh | nh | P | R | T |
| Kappa | nh | A | S | nh | R,L | nh | S | Q | Q | S,Y | nh | S,D | nh | P | R,L | T |

b)

| VH CDR1,2 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| WO1 | G | Y | T | F | T | E | Y | ***T*** | M | H | G | I | ***N*** | P | N | ***N*** | G | T | **S** |
| WO2 | G | Y | S | F | T | G | Y | ***T*** | M | N | L | I | ***N*** | P | Y | ***N*** | G | T | **S** |
| α-Lyso | G | Y | T,S | F | T | T | Y | W | I,V | E,S | E | I | L | P | G | S | nh | T | Y,D |
| Germ | G | F | T | F | T | D | Y | Y | M | S,H | F | I | R | nh | nh | nh | nh | T | E,K |
| Kappa | G | F,Y | T | F | T | D | Y | Y | I,M | H,K | nh | I | nh | P | nh | nh | G,S | T | nh |

| CDR3 | 96 | 97 | 98 | 99 | 100 | 101 | 102 | 103 | 104 | 105 | 106 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| WO1 | C | A | R | D | D | G | - | ***Y*** | Y | G | F |
| WO2 | C | A | R | R | L | G | D | ***Y*** | Y | A | M |
| α-Lys | C | A | R,S | nh | D,E | G | - | nh | Y | G | nh |
| Germl | C | A | R,S | D | nh | nh | nh | nh | Y | A,Y | M,G |
| Kappa | C | A | nh | G,D | nh | G | nh | nh | nh | nh | F |

**Figure 2.** CDR composition of a) $V_L$s and b) $V_H$s compared with those of anti-lysozyme, germline, and kappa-light-chain antibodies. **nh* signifies no homology. Bold, italicized text highlights conserved residues in WO1 and WO2 White-on-black text indicates identical or homologous residues for WO1, WO2, and one or more compared sequences

### 4.1.2 Light Chain

A preliminary sequence alignment was performed using a database of various antibody sequences, including subsets of anti-lysozyme Ab chains and germline Ab chains. The results showed a high degree of homology between heavy and light chain sequences, with several residues noted as distinctive in **Table 1**. The CDRs of the light and heavy chains of WO1 and WO2 are rich in hydrogen-bonding residues relative to anti-lysozyme and germline antibodies, suggesting that hydrogen bonding plays an important role in amyloid-recognition.

A more extensive alignment was done using the Kabat sequence database testing program. The WO1 $V_H$ and $V_L$ sequences were compared to the database of 2707 light chains and 3471 heavy chains. The unique residues Lys42 and Lys74 had no consensus within the database; at these positions, which are not at the CDR, Lys occurred in less than 1% of the sequences. Lys52 at light chain CDR2, however, was a hydroxyl residue in the majority of the 2701 chains. None of the WO2-unique charged or hydroxyl residues showed any incongruity.

| Light Chain Residue WO1 WO2 | Equivalents | Residue exhibited by chains with non-identity |
|---|---|---|
| Lys42 Ser43 | No | None charged, only Gly, Gln, Tyr |
| Lys52 Ser53 | Lys only in WO1 | No charged, 95% Tyr or Ser |
| Asp56 Ser57 | Asp in WO1 and (AL) | 99% Pro, (G) have Thr/Ser |
| Lys74 Thr75 | Lys only inWO1 | 100% Glu/Asp, (G) has Thr |

**Table 1:** Light chain-unique residues and comparison for WO1 residues that are less than 10% homologous to database sequences in the initial alignment tests. The alignment was done with specific interest in the anti-lysozyme Abs as well as germline Abs, and any residues here are noted if appearing in either chain. The residue is listed with its location, as well as the residues that were most commonly shared throughout the other Ab sequences. The equivalent residue from WO2 is also included for comparison. Codes: antilysozyme (AL), germline (G).

### 4.1.3 Heavy Chain

Through the preliminary alignment including the anti-lysozyme, germline, and kappa-light chain sequences, several residues were noted as distinctive. They are listed in **Table 2**. As with the light chain results, the most striking disparities between the WO1/WO2 heavy chains and comparison germline and anti-lysozyme sequences occur with charged residues. Again, the Kabat alignment was performed. The equivalent positions for residues Lys63, Lys65, Lys67, Lys74, Asp73 (in WO1 and WO2) and Arg84 (in WO1 only) have no consensus, but these charged residues are not unique and do not occupy the CDR. The only significant discrepancy was the Asn44 of the WO2 heavy chain. The Asn residue found at this site is found in only 13 examples, 0.326% of the database. Typically, a Ser residue is at this position. This position is also not at the CDR.

| Heavy Chain Residue WO1 WO2 | Equivalents | Residue exhibited by chains with non-identity |
|---|---|---|
| Lys63 Lys63 | WO1/WO2, (G) | No other charged residues |
| Lys65 Lys65 | WO1/WO2, (G), (AL) | 100% Val, Leu, or Phe |
| Lys67 Lys67 | WO1/WO2, (G), (AL) | 100% Gly, Ser, Asp |
| Asp73 Asp73 | WO1/WO2, (G), (AL) | Mostly Asp/Glu |
| Lys74 Lys74 | WO1/WO2, (G) | 99% Asp/Glu |
| Arg84 Leu84 | Arg only in WO1 | Non-consensus, but no other charged |

**Table 2:** Heavy chain unique residues and comparison. The residues displaying less than 10% homology in the initial alignments to database sequences are included here. Any germline or anti-lysozyme Ab commonalities are noted, as well as the residues which occur in the chains which are not homologous to the WO1/WO2 heavy chains. Codes: antilysozyme (AL), germline (G).

### 4.1.4 Fv Models

Of the WO1 light chain's unique lysine residues (42, 52, and 74), only Lys52 is at the CDR and may play a role in docking. However WO1 light-chain Asp56 (unique to WO1 and anti-lysozyme Abs) may play a role, since it makes a hydrogen bond to an Asn residue in the trigonal prism model.

### 4.2 Electrostatic Surface Potentials for WO1 and WO2

WO1 and WO2 Fvs display regions of significant charge density (as calculated by the DelPhi module of InsightII), observed in the electrostatic potential map shown in Figure 3. Though the pIs of the WO1 and WO2 Fvs are nearly neutral, they clearly have highly charged regions. The presence and distribution of these charged segments indicates a possible role of charge interactions in antigen binding. The line of positively charged residues on each of the light-chains suggests a template for hydrogen bonding and salt bridges. The line of negatively charged residues at the heavy-light interface of WO1's CDR (as seen in **Figure 3**) reflects the higher proportion of negatively charged residues in its composition.

### 4.3 Electrostatic Surface Potentials for the Protofilament Models

The most striking features of both models are the bands of charged residues aligned along the stacking axis; these representations are shown in **Figure 4**. The charge bands are a consequence of the in-register parallel-stacking of the models. Labeling the trigonal prism protofilament model's face containing residues 17-22 "A", 23-27 "B", and 28-36 "C" (Aβ 1-40 numbering basis), we see that faces A and C display exposed hydrophobic residues. In the two-strand model, the N-terminal face consists of residues 9-23; this corresponds to the "A" face of the trigonal prism

model. There is then a 180º turn consisting of residues 24-27; this turn (instead of a β sheet) is in sharp contrast to the "B" face of the other model. Finally, residues 30-40 form the C-terminal face of the two-strand model.

**(a)** **(b)**

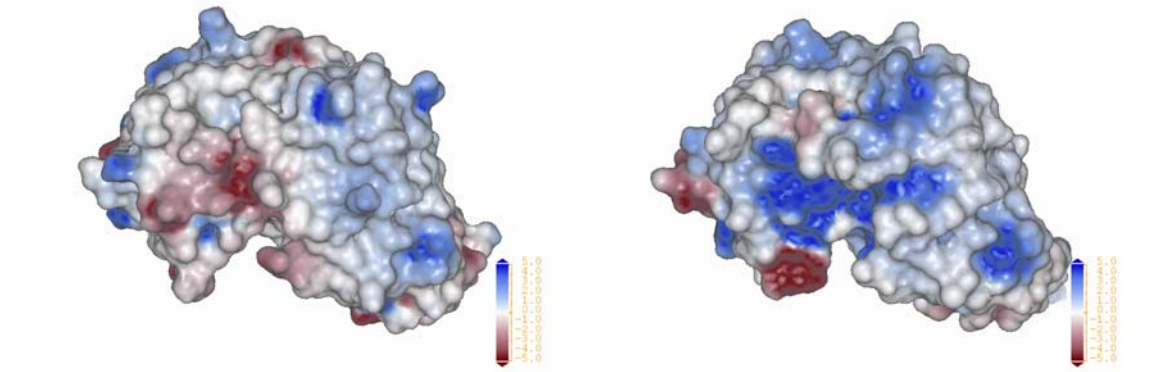


**Figure 3**. Electrostatic surfaces for WO1 **(a)** and WO2 **(b)** Fv models. The viewer faces the CDRs. In each image, the heavy chain fragment is on the left.

**(a)**

**(b)**

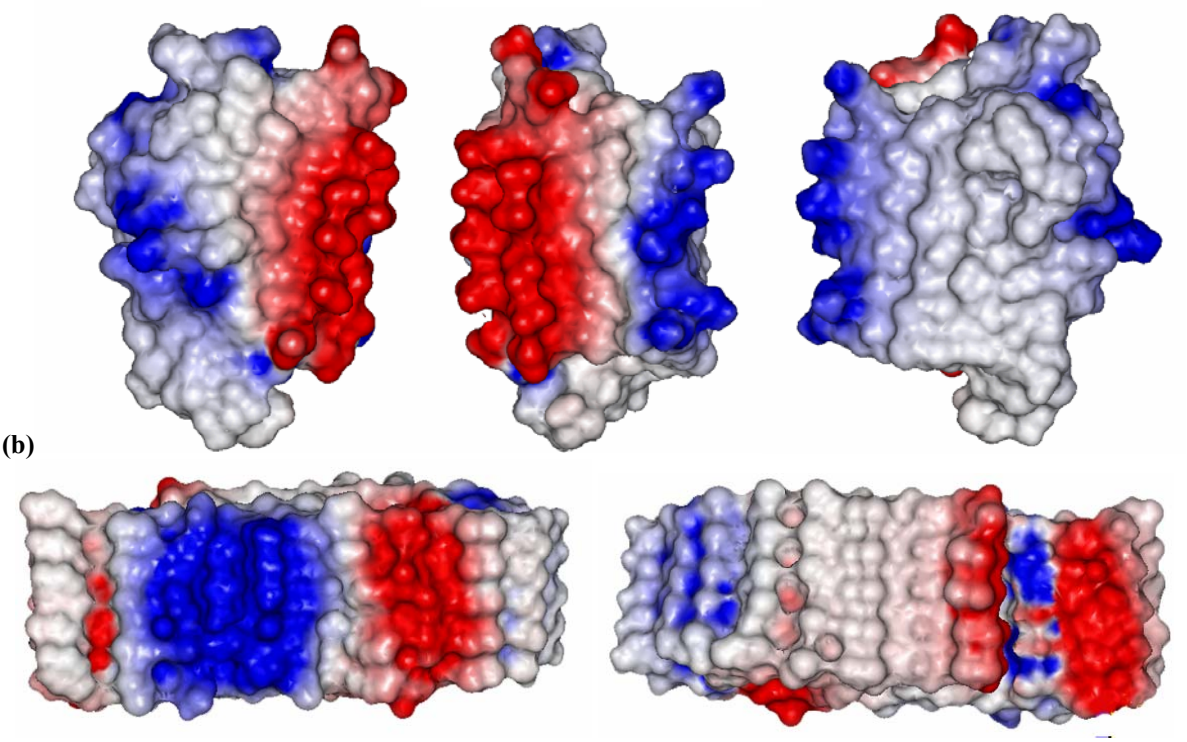


**Figure 4**. **(a)** Electrostatic surfaces for the A, B, and C faces of the trigonal prism protofilament model. The line of Glu22 and Asp23 residues (shown in red) at the corner of the A and B faces is particularly striking. **(b)** Electrostatic surfaces for the two faces of the two-strand model. The N-terminal face is shown in the left-hand image; strand direction is left to right. The C-terminal face is shown in the right-hand image; strand direction is left to right.

### 4.4 Docking WO1-Trigonal Prism-Model

AutoDock3 predicted 50 binding configurations. Both the Fv and the trigonal prism model were treated as rigid bodies. Autodock3 found 41 distinct conformational clusters (using an RMSD-tolerance of 1.0 Å), of which 8 contained more than one solution. The variation in energy from the best to the worst solution was 7.0 kcal/mol.

Although there was one cluster of two solutions in the top 7 (with the B face laterally at the heavy-light interface), more intriguing is a collection of related clusters beginning at the 8th-ranked solution. In total, 21 solutions have the B face of the trigonal prism protofilament model longitudinally oriented at the light-chain CDR, with some overlap onto heavy chain CDR. Some are offset by one, two, or three monomers (i.e., the stack of six Aβ monomers is shifted along the stacking axis), others by angles less than 27˚. This is the largest supercluster; its lowest-energy solution is presented in **Figure 5a**. The particular solution used for the following analysis was the lowest-energy member of this supercluster, which is 3.5 kcal/mol (~1 hydrogen bond) less favorable than the lowest-energy solution of all docking runs.

### 4.4.1 Salt Bridges and Ion Pairs

This docking model indicates that the line of Glu22 and Asp23 residues in the trigonal prism model binds to the corresponding line of polar and positively charged residues of WO1's light-chain CDR through salt bridges and type 2 hydrogen bonds between acidic residues and carbonyl groups. The line of Lys28 residues likewise interact with the trail of polar and negatively charged residues of the light *and* heavy chain CDRs. Breaking the ion pairs down by Szilyagi and Zavodsky classification,[65] there are four bonds at < 4.0 Å, six bonds between 4.0 and 6.0 Å, and ten pairs between 6.0 and 8.0 Å. *Hydrogen bonding*. The O and N atoms of the side chains of the line of Asn27 residues on the three-strand model make hydrogen bonds with Tyr and Asp residues of WO1. The carbonyl O atoms of the line of (inward-pointing) Val24 and Gly25 residues bond to Tyr residues of WO1. The backbone N of inward-pointing Ser26 makes some possible hydrogen-bonds. There are 17 likely hydrogen bonds at less than 3.5 Å, two of which were also counted as close ion pairs. The hydrogen bonding interface is shown in **Figure 5d**.

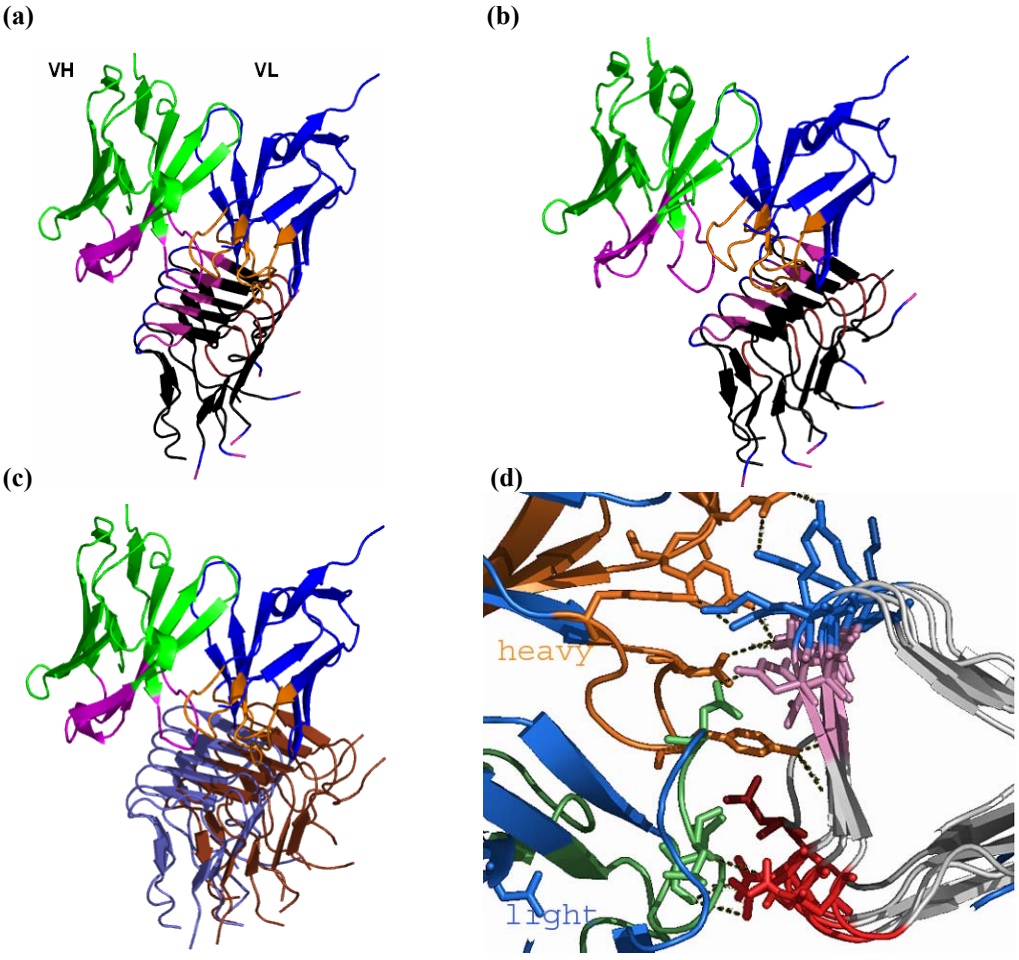



**Figure 5**. **(a)** WO1 and **(b)** WO2 Fv docked with trigonal prism protofilament model in the predicted configuration (top solution in supercluster). The heavy chain is on the left, the light chain is on the right, and the CDRs are highlighted. The trigonal prism model is shown colored according to residue: negative, red; positive, blue; polar, pink; hydrophobic, black. **(c)** The two solutions are showed overlain. **(d)** A close-up view of the docking interface for the WO1-trigonal-prism model. The color code is as for **(a)** and **(b)**, but hydrophobic residues are shown in grey and hydrogen bonds are shown in black.

**4.4.2 Hydrophobic Interactions**
There are no aromatic rings on the B-face of the trigonal prism model, and the hydrophobic Val24 side chain points toward the interior of the trigonal prism model. Thus hydrophobic interactions between WO1 and the trigonal prism model are not observed in this docking model.

**4.5 Docking WO2-Trigonal Prism-Model**
AutoDock3 predicted 50 binding solutions. The Fv and the trigonal prism model were both treated as rigid bodies. For WO2 and the trigonal prism protofilament model, Autodock3 found 42 distinct conformational clusters (using an RMSD-tolerance of 1.0 Å), of which 5 contained more than one member. The variation in energy from the best to the worst solution was 9.3 kcal/mol. There was one cluster of two solutions in the top 13, but the largest cluster began with the $14^{th}$-ranked solution. Within 1.0 Å R.M.S.D., there are 5 similar solutions. In total, 11 solutions have the B face of the trigonal protofilament model at the light-chain CDR, with slight overlap onto heavy chain CDR. One of them is offset by one monomer, others by acute angles. This is the largest supercluster; its lowest-energy solution is shown in **Figure 5b**. This solution resembles the supercluster solution found in the WO1 docking, but makes a 16° angle to it. The particular solution used for the following analysis was the lowest-energy member of this supercluster, which is 6.4 kcal/mol less favorable than the lowest-energy solution.

**4.5.1 Salt Bridges and Ion Pairs**
The analysis for this docked conformation is very similar to that of the WO1-Aβ6 solution, but there are no salt bridges at less than 4.0 Å, only two ion pairs between 4.0 and 6.0 Å, and five between 6.0 and 8.0 Å. *Hydrogen Bonding and hydrophobic interactions.* Thirteen likely hydrogen-bonds at less than 3.5 Å exist between WO2's Fv model and the docked trigonal-prism model. No hydrophobic interactions were observed for this docking model.

**4.6 Docking. WO1-Two-Strand-Model**
AutoDock3 predicted 50 binding solutions. Both the Fv and the protofilament model were treated as rigid bodies. For WO1 and the two-strand protofilament model, Autodock3 found 4 distinct conformational clusters with more than two members (using an RMSD-tolerance of 2.0 Å. The variation in energy from the best to the worst solution was 5.6 kcal/mol. In total, 4 solutions have the N-terminal face on the CDR, 4 have the open end of the two-strand model sitting longitudinally on the CDR, 7 solutions have the "bottom" end (right-hand-rule) of the two-strand protofilament model laterally across the CDR, and 3 have the closed end of the two-strand model near the CDR. The several clusters of near-equal energy suggest a negative result; they are shown in **Figure 6a**. Two of these clusters are not predicted to make any salt bridges at < 4 Å, but the other two are expected to make one and two such bonds, respectively.

**4.7 Docking. WO2-Two-Strand-Model**
As before, AutoDock3 predicted 50 binding solutions with both components (the Fv and the protofilament model) treated as rigid bodies. For WO2 and the two-strand protofilament model, Autodock3 found 35 distinct conformational clusters (using an RMSD-tolerance of 2.0 Å), of which 3 contained more than two members. The variation in energy from the best to the worst solution was 8.6 kcal/mol. The top 7 solutions have the "bottom" end (right-hand-rule) of the two-strand protofilament model laterally across the CDR. This cluster averages 3.8 kcal/mol lower in energy than the next best solution and it is the largest cluster, indicating a positive result; it is displayed in **Figure 6b**. This somewhat resembles the 7-member cluster for WO1-two-strand-model (shown in **Figure 6a**). The particular solution used for the following analysis was the lowest-energy member of the cluster.

**4.7.1 Salt Bridges, Ion Pairs, Hydrogen Bonding, and Hydrophobic Interactions**
No salt bridges were observed out to 4.0 Å, but 10 ion pairs in the range of 4-6 Å suggest a modest electrostatic contribution to WO2 binding Aβ. WO2 and the two-strand model make 7 likely hydrogen bonds at less than 3.4 Å. No *pi*-overlaps are observed.

**4.8 Docking Controls**
Positive controls #1 and #2 used rigid models, positive control #3 allowed flexibility in the ligand. Positive control #1 successfully recreated the crystal structure 1DQJ,[66] lysozyme complexed to antilysozyme. Two of the 50 solutions clustered near that from the crystal structure. These two averaged 11.8 kcal/mol lower in energy than the next most energetically favorable solution. Positive control #2, mouse renin complexed to the angiotensin analog CH-66 (PDB code 1SMR),[67] showed a similar result, with the 44-member dominant cluster 19.7 kcal/mol lower in energy. Positive control #3 did not precisely recreate the crystal structure 1DYI,[68] DHFR complexed with

folate, but rather predicted 10 similar solutions for the 32-atom folate molecule in the correct binding pocket within 2.3 to 10.8 Å RMSD of the true crystal structure, and within 3.5 kcal/mol of one another. Negative control #1 consisted of "docking" the trigonal prism protofilament model with the crystal structure of mouse renin. The result consisted of three superclusters (23, 14, and 5 members) of solutions whose lowest-energy members were within 5.0 kcal/mol of each other. Negative control #2 consisted of "docking" the WO1 Fv model with the inhibitor from the mouse renin crystal structure. All fifty top solutions were within 4 kcal/mol of one another. There was one two-member cluster and one four-member cluster. Solutions ranged all over the Fv.

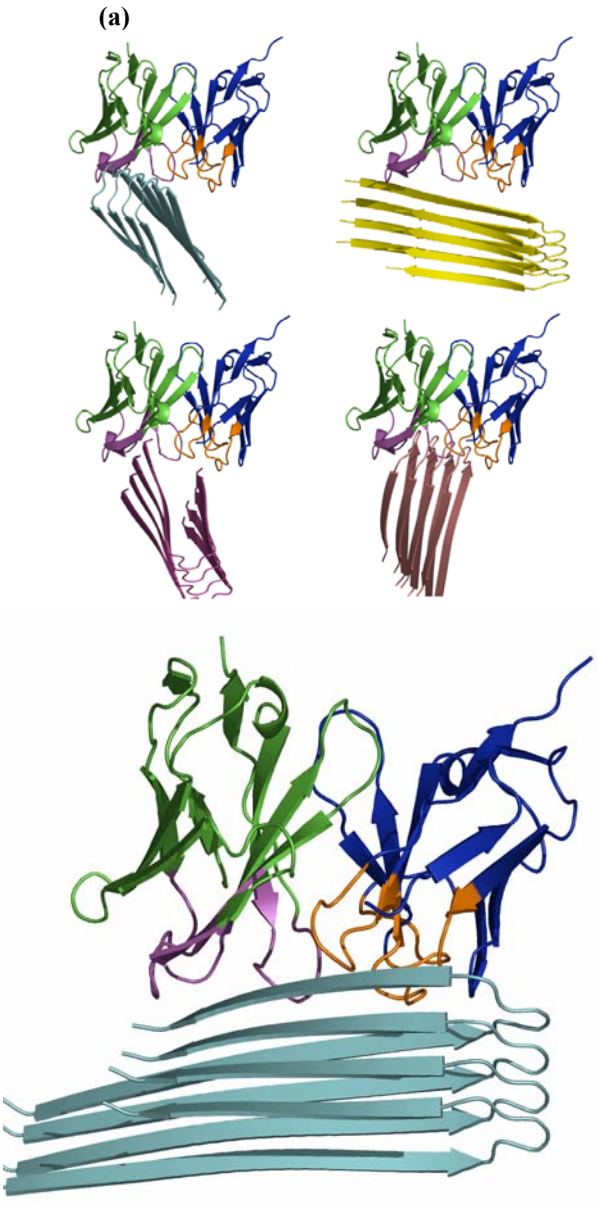



**Figure 6.** **(a)** WO1 Fv docked with two-strand protofilament model in the four predicted configurations. **(b)** WO2 Fv docked with two-strand protofilament model in the top predicted configuration

### 4.9 Binding vs. Salt Concentration

Binding constants were calculated from the sigmoid midpoints of the graph in **Figure 7**. The strongest binding of WO1 and Aβ was observed in PBS; the $EC_{50}$ was 2 nM. Of the 5 mM Hepes conditions, the ones at 150 mM NaCl (nearly-physiological) and at no salt resulted in the strongest binding of WO1 and Aβ, 4 nM. In the presence of 600

mM NaCl, the $EC_{50}$ constant changed sharply to 20 nM, indicating five-fold less binding. Moreover, the 150 mM KCl and 75 mM $Na_xH_yPO_4$ conditions resulted in 8 nM and 10 nM EC50's, respectively. The significantly reduced binding in high salt suggests that electrostatic interactions are relatively important to binding and hydrophobic interactions are relatively unimportant.

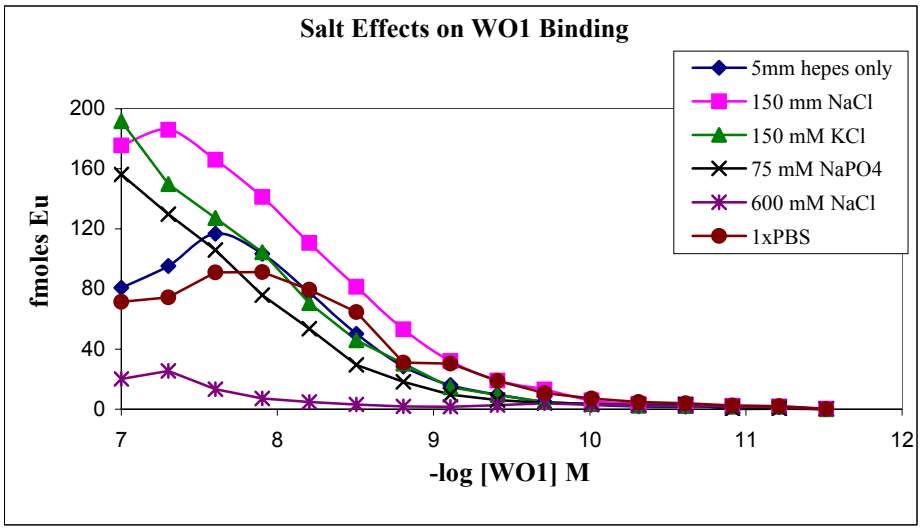



**Figure 7. Salt effect on WO1 binding.** Binding under various salt conditions was calculated from sigmoid midpoints. The $EC_{50}$ values are given in the text.

WO1 binding to Aβ was tested in a range of pH conditions. The Citrate pH 3.0 condition resulted in essentially no binding. Citrate and PIPES at pH 5.8 showed $EC_{50}$ values of 2 nM and >20 nM, respectively. The $EC_{50}$ values for PIPES and Bicine at pH 7.4 were 2 nM and 1.5 nM. For Bicine and CHES at pH 8.5, $EC_{50}$ values were 16 nM and 10 nM. For CHES at pH 10.0, the $EC_{50}$ was > 50 nM. (**Figure 8**). The data suggest that WO1 binds Aβ amyloid best closer to physiological pH values. If hydrophobic interactions dominated WO1- Aβ amyloid binding, this would likely not be the case. This supports the salt effect results suggesting that binding depends at least in part on electrostatic interactions.

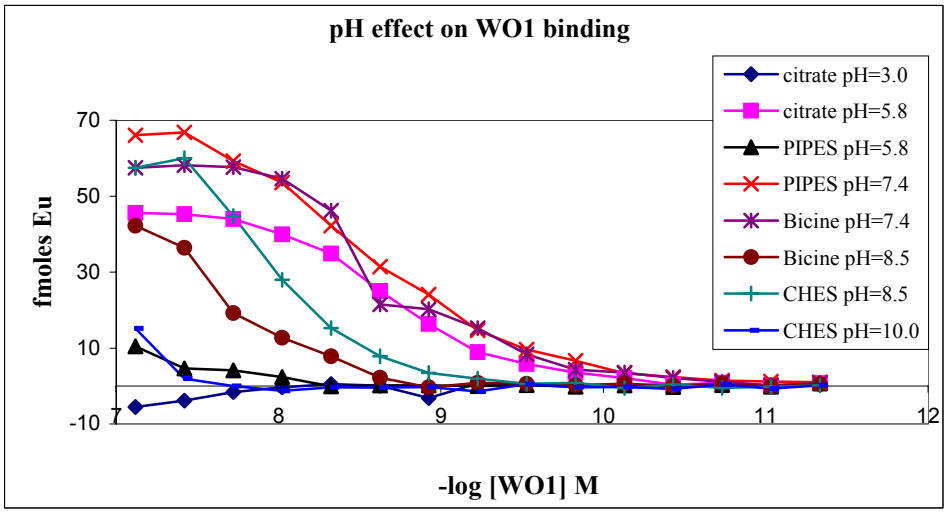



**Figure 8. pH effect on WO1 binding.** The show binding under various pH conditions. Binding affinities were calculated from the sigmoid midpoints. The $EC_{50}$ values are given in the text.

## 5. DISCUSSION

### 5.1 Comparison with Experimental Results

The model of WO1- trigonal prism-model docking that we have proposed, shown in **Figure 5a**, features many hydrogen-bonding pairs and salt-bridges; that agrees with the experimental salt-effect data for WO1- Aβ binding. However, the salt-binding data itself is at odds with WO1's general amyloid-recognition behavior which includes binding to (uncharged) polyglutamine fibrils. This prediction is also called into question by recent data from hydrogen exchange NMR[39] and cysteine scanning mutagenesis[40]. These data are not straightforward to interpret, but they suggest that residues 23-27 (the so-called "B face" of the trigonal-prism protofilament model) may not be in ordered β-sheet structure. There are no strong salt bridges at less than 4.0 Å predicted for binding of WO2 to either the trigonal prism model or the two-strand model.

### 5.2 Assemblies

The Fv models presented here are docked only to a protofilament not to the full Aβ fibril, but these docked models are consistent with the hierarchical fibril model for Aβ amyloid proposed by their authors.[26,38] The fibril model based on the trigonal prism protofilament construction consists of six trigonal prisms packed vertically against one another in two rows of three, in which each protofilament face is exposed twice on the outside surface of the fibril model. For the two-strand model, a full protofilament is generated by applying a two-fold symmetry axis along the C-terminal face perpendicular to the stacking axis of the model. In this way, the open ends of both stacks of monomers point the same way as the stacks twist helically around one another. One depiction of this twisting stack may be found in this work, **Figure 1e**; another is in Petkova, *et al* PNAS 2002, figure 5a. This implies that the C-terminal face of the model is not available for binding; as the predicted docking mode does not employ the C-terminal face, this does not affect our analysis. The Fv model likewise has a single antigen-recognition region for WO1/WO2, whereas the physiological IgM presents ten recognition sites, resulting in high avidity and increased total binding strength.

### 5.3 Implications of Trigonal Prism Results

The docking results for the trigonal prism model suggest that WO1 and WO2 bind to the face of amyloid fibrils similarly. The in-register parallel beta-stacking of the trigonal prism model requires that each residue of the polypeptide line up with its equivalent in the next layer of the prism. Thus, while Glu22 makes peptide bonds to Ala21 and Asp23, it makes hydrogen bonds to Glu22 on neighboring strands. Viewing a face of the prism, then, one would see lines of equivalent residues. The line of Glu22 and Asp23 residues (**Figure 4**) is predicted to bind to the corresponding line of positive and polar residues of WO1's light-chain CDR. The line of Asn27 and Lys28 residues likewise interact with the trail of negative and polar residues.

The lineup of positive and negative charges on the B face with those at the CDR suggests another available binding mode. The A face of the model also has such a lineup of charges (Figure 3). The N-terminal residues of the trigonal prism model occlude part of the A face in this model, which was not allowed torsional freedom, so the A face was less accessible for binding than the B face. Even so, 2/50 solutions for WO1 and 5/50 solutions for WO2, though not clustered, did predict some form of docking to the A face. It is worth noting, furthermore, that earlier simulations with an un-minimized model (in which the N terminal residues took other orientations) yielded a supercluster of docking solutions at the A face of the trigonal prism model with WO1. It is thus likely that a more computationally intensive docking simulation, one which allowed rotation around the bonds of the trigonal prism model, would have shown a bimodal distribution of solutions, one at the A face and another at the B face. These facts, coupled with the new data that suggest non β-sheet conformation for the B-face,[39,40] to some degree weaken the case for the trigonal-prism model for Aβ protofilament. Therefore WO1/WO2 binding to the A face of Aβ remains a distinct possibility. As only 4/50 solutions featured the C face, and they all took very different orientations, our current docking model suggests that the C face is unlikely to feature in WO1 binding.

### 5.4 Implications of Two-Strand Results

Docking calculations based on the two-strand model do not indicate a conclusive result for WO1-binding. However, computations for WO2 support WO2 binding to the end of elongating amyloid fibrils, while only the 48$^{th}$ ranked hit for WO2 and the trigonal prism resembles such an arrangement. This contrasts with expectation that the lines of charge along the N-terminal face would provide a recognition site for the antibodies, but is consistent with the fact that these antibodies bind amyloid fibrils (such as polyglutamine) that lack charged residues. Also, the lower degree

of order found in residues 23-27 in the two-strand model may be more consistent with the new hydrogen exchange and cysteine mutagenesis data.

## 6. CONCLUSIONS

The sequences of two amyloid-recognizing IgMs were compared and analyzed. Structural models of their Fvs were constructed. WO1's Fv was found to be unusually rich in charged residues, while WO2 is rich in Ser residues. As one might expect from the differences in their secondary and ternary structures, the two different protofilament models lead to different predicted docking models. Results of docking simulation for the trigonal-prism model imply that WO1 and WO2 bind to the face of a fibril of Aβ. The results for the two-strand model make no positive prediction for WO1, but results for the two-strand model suggest that WO2 binds to the ends of fibrils. The results for WO1 and the trigonal prism model agree with experimental salt-binding data, while the results for WO2 and the two-strand model are more consistent with WO1/WO2's diverse binding behavior. While these results do not support either model over the other, they make testable predictions.

## 7. ACKNOWLEDGMENTS

We thank Doctors Juntao Guo and Ying Xu for the trigonal prism protofilament model and Doctor Robert Tycko for the two-strand protofilament model.

## 8. REFERENCES

[1] R. H. Falk, R. L. Comenzo and M. Skinner, *N Engl J Med* 337 (1997).
[2] D. J. Selkoe, *Ann N Y Acad Sci* 924 (2000), 17.
[3] I. Tariska, *Acta Med Acad Sci Hung* 21 (1965), 483.
[4] P. A. Temussi, L. Masino and A. Pastore, *Embo J* 22 (2003), 355.
[5] E. T. Jaikaran and A. Clark, *Biochim Biophys Acta* 1537 (2001), 179.
[6] M. R. Hayden and S. C. Tyagi, *Jop* 2 (2001), 124.
[7] O. Corti and A. Brice, *J Soc Biol* 196 (2002), 95.
[8] R. Kruger, O. Eberhardt, O. Riess and J. B. Schulz, *Trends Mol Med* 8 (2002), 236.
[9] J. M. George, *Genome Biol* 3 (2002), REVIEWS3002.
[10] M. Pras, F. Prelli, E. C. Franklin and B. Frangione, *Proc Natl Acad Sci U S A* 80 (1983), 539.
[11] H. D. Feiner, *Hum Pathol* 19 (1988), 1255.
[12] J. N. Buxbaum, J. V. Chuba, G. C. Hellman, A. Solomon and G. R. Gallo, *Ann Intern Med* 112 (1990), 455.
[13] V. Bellotti, P. Mangione and G. Merlini, *J Struct Biol* 130 (2000), 280.
[14] C. A. Ross, *Neuron* 35 (2002), 819.
[15] C. Weissmann, *Nat Rev Microbiol* 2 (2004), 861.
[16] H. Inouye, P. E. Fraser and D. A. Kirschner, *Biophys J* 64 (1993), 502.
[17] M. Sunde and C. Blake, *Adv Protein Chem* 50 (1997), 123.
[18] S. B. Malinchik, H. Inouye, K. E. Szumowski and D. A. Kirschner, *Biophys J* 74 (1998), 537.
[19] A. D. Williams, et al., *J Mol Biol* 335 (2004), 833.
[20] J. D. Harper, S. S. Wong, C. M. Lieber and P. T. Lansbury, Jr., *Biochemistry* 38 (1999), 8972.
[21] L. C. Serpell, et al., *J Mol Biol* 300 (2000), 1033.
[22] T. C. Saido, W. Yamao-Harigaya, T. Iwatsubo and S. Kawashima, *Neurosci Lett* 215 (1996), 173.
[23] I. Kheterpal, A. Williams, C. Murphy, B. Bledsoe and R. Wetzel, *Biochemistry* 40 (2001), 11757.
[24] O. N. Antzutkin, R. D. Leapman, J. J. Balbach and R. Tycko, *Biochemistry* 41 (2002), 15436.
[25] M. Torok, et al., *J Biol Chem* 277 (2002), 40810.
[26] A. T. Petkova, et al., *Proc Natl Acad Sci U S A* 99 (2002), 16742.
[27] T. S. Burkoth, et al., *J. Am. Chem. Soc.* 122 (2000), 7883.
[28] I. Kheterpal, et al., *Biochemistry* 42 (2003), 14092.
[29] S. S. Wang, S. A. Tobler, T. A. Good and E. J. Fernandez, *Biochemistry* 42 (2003), 9507.
[30] M. Kraus, M. Bienert and E. Krause, *Rapid Commun Mass Spectrom* 17 (2003), 222.
[31] I. Kheterpal, S. Zhou, K. D. Cook and R. Wetzel, *Proc Natl Acad Sci U S A* 97 (2000), 13597.
[32] T. L. Benzinger, et al., *Proc Natl Acad Sci U S A* 95 (1998), 13407.
[33] M. F. Perutz, J. T. Finch, J. Berriman and A. Lesk, *Proc Natl Acad Sci U S A* 99 (2002), 5591.
[34] L. O. Tjernberg, et al., *Biochem J* 366 (2002), 343.

[35] C. Hilbich, B. Kisters-Woike, J. Reed, C. L. Masters and K. Beyreuther, *J Mol Biol* 218 (1991), 149.
[36] M. O. Chaney, S. D. Webster, Y. M. Kuo and A. E. Roher, *Protein Eng* 11 (1998), 761.
[37] L. C. Serpell, C. C. Blake and P. E. Fraser, *Biochemistry* 39 (2000), 13269.
[38] J. T. Guo, R. Wetzel and Y. Xu, *Proteins* 57 (2004), 357.
[39] N. A. Whittemore, et al., *Biochemistry* 44 (2005), 4434.
[40] S. Shivaprasad and R. Wetzel, *Biochemistry* 43 (2004), 15310.
[41] R. Nelson, et al., *Nature* 435 (2005), 773.
[42] B. O'Nuallain and R. Wetzel, *Proc Natl Acad Sci U S A* 99 (2002), 1485.
[43] M. Dumoulin and C. M. Dobson, *Biochimie* 86 (2004), 589.
[44] A. C. Davis and M. J. Shulman, *Immunol Today* 10 (1989), 118.
[45] N. R. Whitelegg and A. R. Rees, *Protein Eng* 13 (2000), 819.
[46] Chemical Computing Group Inc., 1997.
[47] E. E. Abola, J. L. Sussman, J. Prilusky and N. O. Manning, *Methods Enzymol* 277 (1997), 556.
[48] J. L. Sussman, et al., *Acta Crystallogr D Biol Crystallogr* 54 (1998), 1078.
[49] A. C. Martin, *Proteins* 25 (1996), 130.
[50] S. F. Altschul, et al., *Nucleic Acids Res* 25 (1997), 3389.
[51] C. Chothia and A. M. Lesk, *J Mol Biol* 196 (1987), 901.
[52] T. T. Wu and E. A. Kabat, *Mol Immunol* 29 (1992), 1141.
[53] A. C. Martin, J. C. Cheetham and A. R. Rees, *Proc Natl Acad Sci U S A* 86 (1989), 9268.
[54] A. C. Martin, J. C. Cheetham and A. R. Rees, *Methods Enzymol* 203 (1991), 121.
[55] A. R. Rees, Searle,S.M.J., Henry,A.H.and Pedersen,J.T., *in* "Protein Structure Prediction" (S. M.J.E., ed.), p. 141–172. Oxford University Press, Oxford, 1996.
[56] J. T. Pedersen, Searle,S.M.J., Henry,A.H. and Rees,A.R., *Immunomethods* 1 (1992), 126.
[57] G. M. Morris, et al., *Journal of Computational Chemistry* 19 (1998), 1639.
[58] S. J. K. Weiner, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P., *J Am Chem Soc* 106 (1984), 765.
[59] S. J. K. Weiner, P. A.; Nguyen, D. T.; Case, D. A., *J Comput Chem* 7 (1986), 230
[60] Accelrys; MSI, San Diego, 2000.
[61] G. Murshudov, *Acta Crystallogr D Biol Crystallogr.* 53 (1997), 240.
[62] N. Collaborative Computational Project, *Acta Crystallogr D* 50 (1994), 760.
[63] C. Cole and J. Warwicker, *Protein Sci* 11 (2002), 2860.
[64] W. L. DeLano. DeLano Scientific, San Carlos, CA, USA, 2002.
[65] A. Szilagyi and P. Zavodszky, *Structure Fold Des* 8 (2000), 493.
[66] Y. Li, H. Li, S. J. Smith-Gill and R. A. Mariuzza, *Biochemistry* 39 (2000), 6296.
[67] C. G. Dealwis, et al., *J Mol Biol* 236 (1994), 342.
[68] V. M. Reyes, M. R. Sawaya, K. A. Brown and J. Kraut, *Biochemistry* 34 (1995), 2710.

# SNAPSHOTS OF THE PANTOTHENATE SYNTHETASE FROM MYCOBACTERIUM TUBERCULOSIS ALONG THE REACTION COORDINATE

Shuishu Wang
*Public Health Research Institute, 225 Warren Street, Newark, New Jersey 07103*

David Eisenberg
*Howard Hughes Medical Institute, UCLA-DOE Institute of Genomics and Proteomics, Molecular Biology Institute, UCLA, Box 951570, Los Angeles, CA 90095-1570*

## 1. ABSTRACT

Pantothenate synthetase (PS) from *Mycobacterium tuberculosis* represents a potential target for developing new anti-tuberculosis drugs. PS catalyzes the ATP-dependent condensation of pantoate and β-alanine to form pantothenate. We determined the crystal structures of PS from *M. tuberculosis* and its complexes with substrates ATP, pantoate, and β-alanine, as well as an ATP analog AMPCPP, a reaction intermediate pantoyl adenylate and a reaction product AMP, with resolutions from 1.6 to 2 Å. The PS structure reveals a dimer, and each subunit has two domains with tight association between domains. The active site cavity is on the N-terminal domain, covered by the C-terminal domain. There is a flexible loop that forms one wall of the active site cavity, and it opens and closes the active site cavity. The enzyme binds ATP and pantoate tightly in the active site, and brings the carboxyl oxygen of pantoate near the α-phosphorus atom of ATP for an in-line nucleophilic attack. The pantoyl adenylate thus formed is stabilized by tight binding interactions with enzyme active site residues, and protected from hydrolysis by closing of the active site cavity with the flexible loop becoming ordered. Binding of β-alanine can occur only after formation of the pantoyl adenylate intermediate. Crystal structures of these complexes provide a step-by-step view of the PS catalyzed reaction, and allow us to construct transient reaction intermediates and deduce detailed reaction mechanism. PS catalyzes the reaction by stabilizing the reaction intermediates along the reaction coordinate, and structural models of these intermediates can provide basis for inhibitor design.

## 2. INTRODUCTION

Pantothenate (vitamin B5) is an essential precursor for the biosynthesis of coenzyme A and acyl carrier proteins, both of which play critical roles in many cellular processes including energy metabolism and fatty acid metabolism [1]. Microorganisms and plants can synthesize pantothenate, while animals obtain this essential nutrient from their diet [2]. Therefore, the pantothenate biosynthetic pathway offers targets for developing drugs against microbial pathogens. Recently, Jacobs and coauthors [3] reported that a *Mycobacterium tuberculosis* (MTB) mutant defective in the de novo biosynthesis of pantothenate is highly attenuated in both immunocompromised and immunocompetent mice. This observation indicates that a functional pantothenate biosynthetic pathway is essential for virulence of MTB, and thus represents a new target for anti-tuberculosis agents.

The pantothenate biosynthetic pathway in bacteria comprises four steps catalyzed by enzymes encoded by the *panB*, *panC*, *panD* and *panE* genes [4]. The *panC* gene encodes a pantothenate synthetase (PS), which catalyzes the last step of pantothenate biosynthesis, the ATP-dependent condensation of pantoate and β-alanine to form pantothenate. The *panC* gene product in several organisms has been identified and characterized, including those in *Escherichia coli* [5], *Saccharomyces cerevisiae*, higher plants *Oryza sativa* and *Lotus japonicus* [6], fungus *Fusarium oxysporum* [7], and *M. tuberculosis* [8]. The PS enzymes from *E. coli*, higher plants, and MTB form dimers in solution.

The *E. coli* PS enzyme structure belongs to the cytidylyltransferase superfamily [5]. It has two distinct domains, a large N-terminal domain having a Rossmann fold and a smaller C-terminal domain containing a helical layer above a three-stranded antiparallel β-sheet. Based on structural comparison of the *E. coli* PS with other members of cytidylyltransferase superfamily having known structures, von Delft et al. [5] deduced the ATP and pantoate binding sites of the *E. coli* PS and proposed a hinged domain mechanism for opening and closing of the enzyme active site cavity.

Kinetic analysis of the MTB PS suggests that the enzyme-catalyzed reaction proceeds through two steps: the formation of an enzyme bound intermediate, pantoyl adenylate, from ATP and pantoate, followed by nucleophilic attack on the intermediate by β-alanine to form pantothenate and AMP [8]. The existence of pantoyl adenylate as an enzyme bound intermediate is suggested by the transfer of the $^{18}O$ label from the carboxyl group of pantoate to the reaction product AMP. However, pantoyl adenylate cannot be directly isolated or identified by TLC methods, presumably due to rapid lactonization [9].

Here, we present the crystal structure of the MTB PS enzyme and structures of its complexes with substrates, a substrate analog (AMPCPP), a reaction product (AMP), and a pantoyl adenylate reaction intermediate [10, 11]. The MTB PS has the same fold as the *E. coli* enzyme. However, in the MTB PS structure the domains of each subunit have a closed conformation, in contrast to those of the *E. coli* PS structure [5], and there is no significant movement between domains among all structures of the apo-enzyme and various complexes. A flexible region, which forms a wall of the active site cavity, becomes ordered in the reaction intermediate complex and closes the active site cavity, thus acting as a gate to the active site cavity of the MTB PS enzyme. The structure of the β-alanine complex gives direct evidence that the binding site for this last substrate exists only after formation of the pantoyl adenylate intermediate. Together these crystal structures lead to a detailed view of the overall reaction mechanism. Models of the transient reaction intermediate can be easily constructed from the experimentally observed complexes.

## 3. MATERIALS AND METHODS

### 3.1 Protein Production

The detailed procedures for cloning of the *panC* gene, expression and purification of the protein have been described [10]. Briefly, the MTB *panC* gene (Rv3602c) encoding the pantothenate synthetase was amplified from the genomic DNA of MTB strain H37Rv and inserted into a pET30a plasmid (Novagen), which produces a recombinant protein with an N-terminal 6xHis tag that can be cleaved off with enterokinase. BL21(DE3) cells containing the pET30-*panC* plasmid were grown in LB medium containing 50 μg/ml kanamycin at 37 °C and induced with 0.4 mM IPTG for 3 hours. Cells were collected and lysed by lysozyme treatment followed by sonication in 20 mM HEPES pH 7.8, 500 mM NaCl, and 0.5 mM PMSF. The protein was purified from a $Ni^{2+}$-charged HiTrap chelating column (Amersham) with 20 mM HEPES pH 7.8, 500 mM NaCl and a linear gradient of imidazole. The PS protein was subjected to enterokinase (New England Biolabs) digestion to cleave off the N-terminal fusion tag. Electrospray mass spectroscopy indicated that 9 residues from the C-terminus of the protein were also cleaved off by enterokinase digestion. However, enzyme activity assays suggested that this C-terminal truncation does not have any detectable effect on the enzyme activity [10]. The enterokinase digested protein was further purified with a second round of chromatography on a $Ni^{2+}$-HiTrap column followed by Superdex 75 column (Amersham) gel filtration.

### 3.2 Crystallization and Data Collection

Crystallization was carried out as described previously [10]. Crystals in space group $P2_1$ were obtained from drops set up with well solutions containing 10-15% PEG 3000, 5% glycerol, 2% ethanol, 20 mM $MgCl_2$, 150 mM $Li_2SO_4$ and 100 mM imidazole pH 8.0 at 20 ºC. Those in space group C2 were from similar conditions, except with 2% isopropanol in place of ethanol and with 200 mM $Li_2SO_4$. Cocrystallization with β-alanine was done with a well solution of 15% PEG 3000, 2% ethanol, 20 mM $MgCl_2$, 100 mM imidazole pH 8.0 and 20 mM β-alanine, from which crystals in space group $P2_12_12_1$ were obtained. However, these crystals do not have β-alanine in the active site. Cocrystallization with both ATP and pantoate was carried out in the same crystallization condition of the $P2_1$ crystals but with both ATP and pantoate added, from which isomorphous $P2_1$ crystals were obtained that have a pantoyl adenylate intermediate in the active site.

Crystals in complex with substrates were obtained by soaking apo-enzyme crystals in solutions containing substrates or the ATP analog AMPCPP. Soaking experiments were carried out by adding solutions containing substrates or AMPCPP directly to the drops, or to drops of pseudo mother liquor with transferred crystals, and incubating overnight or longer. Crystals were soaked in solutions containing ~10 mM each of AMPCPP alone, both AMPCPP and pantoate, both pantoate and β-alanine, both ATP and pantoate, and both AMP and β-alanine. Crystals were also soaked in solutions containing up to 20 mM pantothenate, but no pantothenate was found in the crystal structure.

Before data collection, crystals were soaked for 2-5 min in a cryogenic solution similar to well solutions or substrate soaking solutions with glycerol added to 30%, and they were then flash-frozen in a cryo stream of $N_2$ gas at 100 K. Diffraction data were collected at 100 K on a Rigaku FRD generator with an R-AXIS $IV^{++}$ detector. Data reduction and scaling were carried out with the programs DENZO and SCALEPACK [12]. Data processing statistics are reported in references [10] and [11]. Table 1 summarizes a few of the data sets, of which the structures are mentioned in this paper. The $P2_1$ crystal has two PS molecules per asymmetric unit; while the C2 crystal contains one molecule per asymmetric unit. In the C2 crystal, the active site is locked in an inactive conformation by the crystal packing (see below), and the pantoate and β-alanine binding sites are compromised. Therefore, only AMP was found when we soaked the C2 crystals with both AMP and β-alanine (data set 8, in Table 1), and those soaked with both ATP and pantoate gave an ATP complex (data set 9) instead of a reaction intermediate complex.

**TABLE 1**

Summary of crystal structure data and experiments

| | Complex | Experiment | SG | #/A.U. | resolution | PDB ID | Ref. |
|---|---|---|---|---|---|---|---|
| 1 | Apo enzyme | | $P2_1$ | 2 | 1.6 Å | 1MOP | [10] |
| 2 | AMPCPP and pantoate in A; AMPCPP in B | AMPCPP, pantoate, soak | $P2_1$ | 2 | 1.6 Å | 1N2E | [10] |
| 3 | Pantoyl adenylate | ATP, pantoate, cocrystal | $P2_1$ | 2 | 2.0 Å | 1N2H | [10] |
| 4 | Pantoyl adenylate | ATP, pantoate, soak | $P2_1$ | 2 | 1.7 Å | 1N2I | [10] |
| 5 | Pantoate | Pantoate, β-alanine, soak | $P2_1$ | 2 | 1.8 Å | 1N2J | [10] |
| 6 | AMP, β-alanine in A; AMP only in B | AMP, β-alanine soak | $P2_1$ | 2 | 1.85 Å | 1A86 | [11] |
| 7 | Apo enzyme | | C2 | 1 | 1.7 Å | 2A88 | [11] |
| 8 | AMP | AMP, β-alanine soak | C2 | 1 | 1.7 Å | 2A7X | [11] |
| 9 | ATP | ATP, pantoate, soak | C2 | 1 | 1.6 Å | 2A84 | [11] |

### 3.3 Structural Determination and Refinement

The crystal structure of the apo-enzyme in the $P2_1$ crystal was determined [10] by the molecular replacement method using AMORE [13], with subunit A of the *E. coli* PS structure (PDB ID code 1IHO [5]) as a model. The two domains were separated as two independent search models. The positions of the two N-terminal domains and one C-terminal domain were found from the rotation and translation searches, and the other C-terminal domain was generated by a non-crystallographic-symmetry (NCS) operation on the position of the known C-terminal domain with the NCS symmetry operators derived from the two N-terminal domains. The structure of the apo-enzyme was used to calculate phases for data sets of various complexes by rigid body refinement into new data sets for isomorphous crystals, or by molecular replacement with AMORE. The final refinement statistics are reported in references [10, 11]. Structural models were refined against diffraction data using CNS [14] and/or REFMAC [15]. After each cycle of refinement, models were manually adjusted with electron density maps using O [16]. The same subsets of data for $R_{free}$ calculation were kept between the CNS and REFMAC programs.

Besides the 9 residues that were digested off by enterokinase, there are 10 more residues from the C-terminus that do not have clear electron density. Based on its susceptibility to protease digestion and lack of clear electron density, it is reasonable to assume that these residues at the C-terminus of the protein are flexible, and do not have a defined structure in solution. It is clear from the structure that both N- and C-termini are away from the active site cavity, and therefore are unlikely to affect the enzyme catalytic activity. This is consistent with the results of enzyme activity assays [10].

## 4. STRUCTURE OF THE MTB PANTOTHENATE SYNTHETASE

### 4.1 Dimer Structure

The crystal structure of the MTB PS enzyme indicates that it is a dimer (**Figure 1**), resembling a butterfly when viewed from one side. The dimer interface is extensive, having a buried area of about 2150 $Å^2$. A cluster of 8 hydrophobic side chains at the center of the dimer interface, Phe174, Leu177, Leu144, and Val118, form the core of the dimer interaction. This cluster is surrounded by hydrogen bonds and salt bridges, as well as the specific β strand

interactions of the two-stranded intersubunit β sheet that is composed of the strand β5 from each subunit. The MTB PS enzyme also exists in solution as a dimer [8, 10], as do PS enzymes from higher plants [6, 7].

The two subunits of the dimer in the $P2_1$ crystal are similar to each other, with an rmsd of 0.9 for all Cα atoms except the disordered residues. Most of the deviation lies at residues 74 to 88, and residues 260 to 265, where the crystal packing environments are different. Residues 74 to 88 (β3-$3_{10}$3-α3'-α3) form one wall of the active site cavity (see **Figure 2** and the active site cavity section b 31 In subunit A these residues are ordered with residues 74 to 76 having high B factors, while in subunit B re- 75 to 80 (helix α3'and nearby residues) do not have electron density, and the rest of the residues up to residue 88 are partially disordered with high B factors. This flexible wall of the active site cavity becomes ordered in the complex with the pantoyl adenylate reaction intermediate, suggesting their function as a gate to the active site cavity (see below pantoyl adenylate binding interactions).

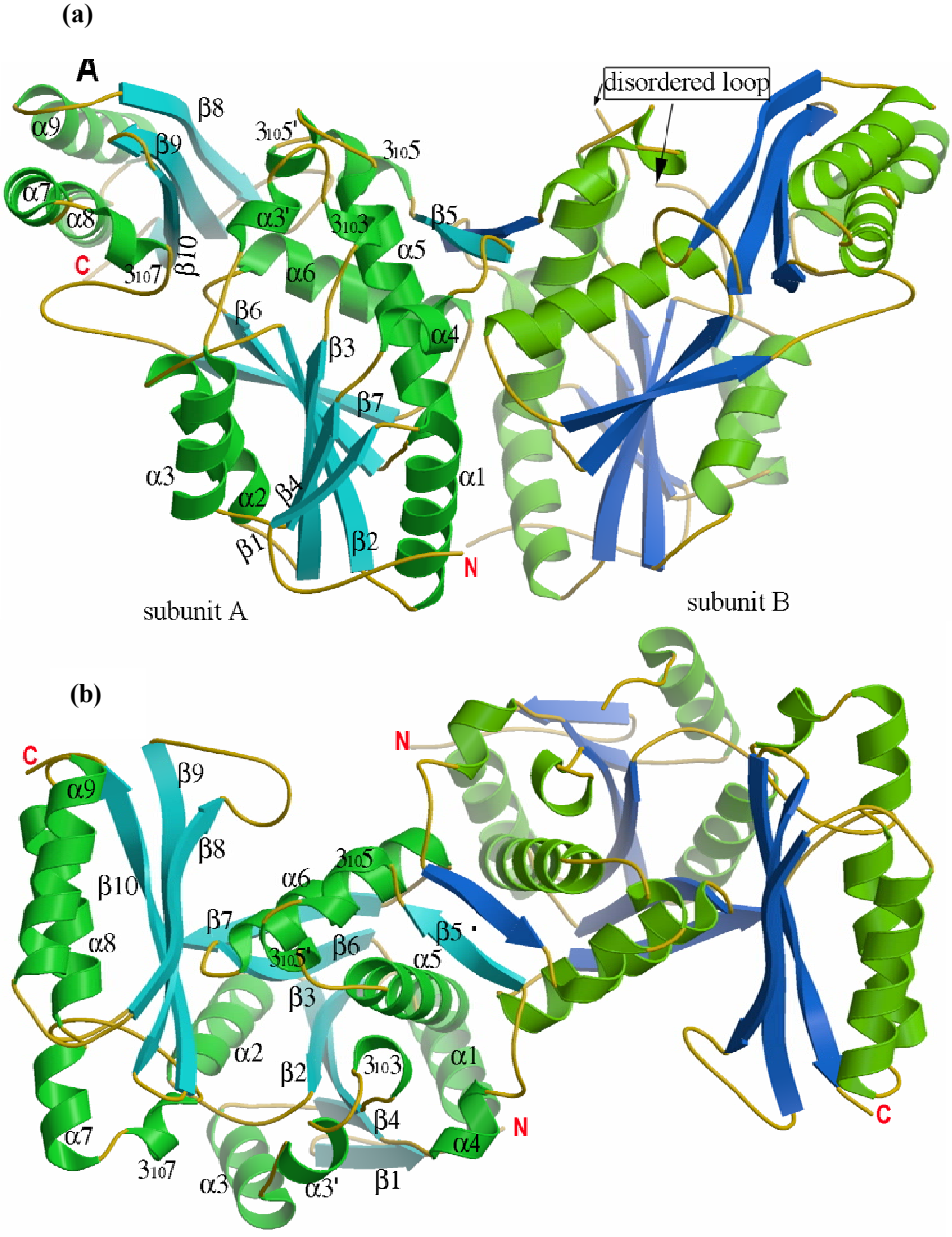



**Figure 1.** Ribbon diagram of the *M. tuberculosis* pantothenate synthetase dimer. **(a)** A side view of the dimer structure shows that it resembles the shape of a butterfly. **(b)** An orthogonal view of (A) from top, with the two-fold NCS symmetry axis (labeled with a dot) approximately perpendicular to the paper plane. Secondary structure elements for subunit A (left) are labeled. Those for subunit B are identical except that the short helix α3' is disordered. The figure was prepared from the apo-enzyme structure of the $P2_1$ crystal, with Molscript [17] and Raster3D [18].

The C2 crystal differs from the $P2_1$ crystal in that there is one polypeptide chain per asymmetric unit [11]. However, the protein molecule in the C2 crystal is also present as a dimer that is essentially identical to that in the $P2_1$ crystal, but the two-fold axis of the dimer coincides with one crystal symmetry axis. In the C2 crystal, a larger portion of the flexible wall, from residue 76 to 86, is completely disordered. The first turn of helix α3 unwinds, and helix $3_{10}3$, which contains residue Glu72 that is important for binding pantoate (see below pantoate binding site), also unwinds and forms an extended structure. This 32 ral difference is induced by crystal packing and the intrinsic flexibility of the loop.

### 4.2 Domain Interactions

Each subunit of the dimer has two well-defined domains (**Figure 1**). The N-terminal domain has a Rossmann fold, with a central parallel β sheet composed of strands β1 to β4, β6, and β7, and helices on both sides of the β-sheet. Strand β7 leads to the C-terminal domain, which starts with a hairpin loop followed by helix $3_{10}7$. The rest of the C-terminal domain has a simple two-layer structure: a layer of helices α7, α8, and α9 followed by a layer of antiparallel β sheet composed of strands β8 to β10.

The two domains have extensive interactions. There is a hydrophobic core composed of residues Leu123, Leu127, Tyr162, Val166 in the N-terminal domain and residues Tyr249, Leu257, and Leu269 in the C-terminal domain. Surrounding this hydrophobic core are three salt bridges, Glu159 to Arg267, Glu128 to Arg278, Glu126 to Arg253, and 6 additional hydrogen bonds linking the two domains. In addition, there are several water-mediated hydrogen bonding interactions. The relative positions of the two domains stay the same for the two subunits in the dimer and for all crystals including the $P2_1$, $P2_12_12_1$, and C2 crystals. These structural features suggest that domain hinge motion is unlikely to be the mechanism of opening and closing the active site cavity for MTB PS.

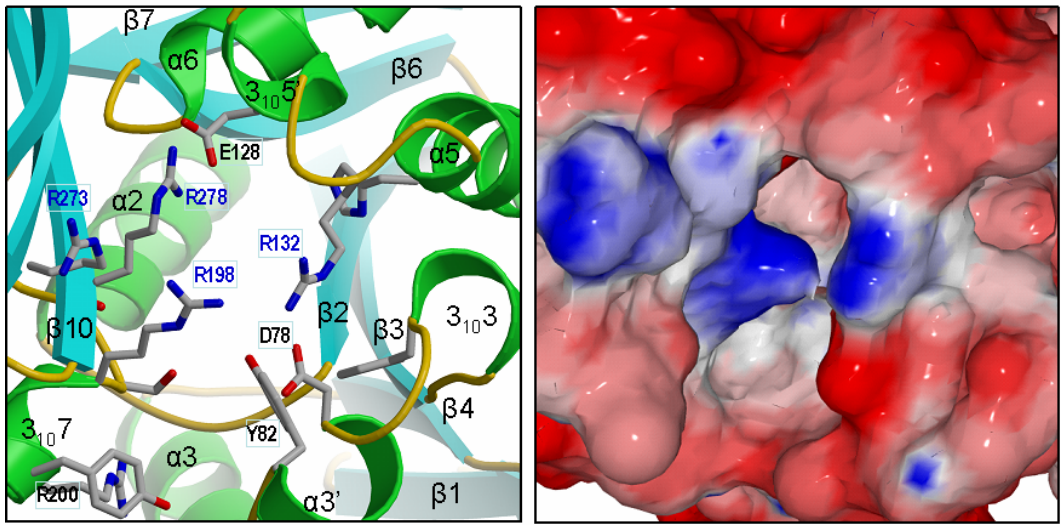



**Figure 2.** Active site cavity and the electrostatic potential surface around the entrance to the active site cavity. The figure was prepared from the subunit A of the $P2_1$ crystal. The two panels are approximately the same view. In the left panel, side chains at the entrance to the activity are shown as sticks. The segment from the end of β3 to the beginning of α3 forms the flexible wall, which is disordered in subunit B and in the C2 crystal thus opening the active site cavity.

### 4.3 Active Site Cavity

Typical for nucleotide binding (Rossmann fold) proteins, the active site is located at the C-terminus of the central parallel β sheet. The cavity is large and deep, and its bottom is at a cleft between strands β2 and β6 (**Figure 2**). The segment between β3 and α3 (loop-$3_{10}$3-loop-α3'-loop region) forms one side of the walls surrounding the cavity in subunit A of the $P2_1$ crystal. Part of this loop is disordered in subunit B, thus opens up the active site cavity. Other walls of the cavity are from β2-loop-α2, $3_{10}5'$-loop-α5, β6-loop-α6, and the loop after β7. Helix $3_{10}7$ and the β strands from the C-terminal domain partially cover the top of the active site cavity. The bottom of the active site cavity is mainly hydrophobic, while the top half of the cavity has several charged residues, including His44 and His47 at the N-terminus of α2 (the HIGH motif [5, 19]), Lys160 and Asp161 at the N-terminus of α6, and Arg198 of $3_{10}7$ (KMSKS motif [5]). At the entrance to the active site cavity, several charged side chains stick to the center,

leaving only a small opening to the cavity (**Figure 2**, right panel). Four arginine side chains, including Arg198, form a patch of positive charges at the entrance. These positive charges may steer the negatively charged substrates into the active site cavity. As we will see later, Arg198 plays an important role in binding the phosphate groups of ATP.

### 4.4 ATP Binding Interactions

The active site is well designed for binding ATP. **Figure 3** shows the initial difference electron density for the ATP molecule in the active site of the C2 crystal. The ATP molecule has an average temperature factor of 17.3 $Å^2$, virtually identical to the average temperature factor of the surrounding protein atoms, indicating that it is at full occupancy. A magnesium ion was found binding to the phosphate groups of ATP. It has a nearly perfect octahedral coordination. Its distances to the O2α, O2β, and O1γ atoms of the phosphate groups are 2.26 Å, 1.99 Å, and 2.12 Å, respectively; three other ligands are water molecules with distances of 1.96 Å, 2.11 Å, and 2.28 Å. The adenosine group of ATP binds tightly at the bottom of the active site cavity with hydrogen bonds and hydrophobic interactions [10, 11]. The phosphate groups stick toward the top of the active site cavity, and are located near the N-terminal ends of helices α2 and $3_{10}7$. Binding of the phosphate groups involves mainly main chain NH groups and charged side chains.

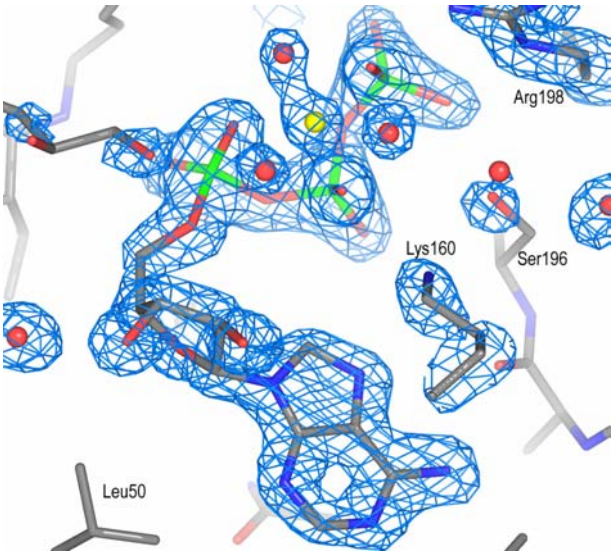



**Figure 3**. Initial Fo – Fc electron density in the active site of the C2 crystal soaked in a solution containing ATP and pantoate superimposed on the final refined model. The map was calculated at 1.6 Å and contoured at 3.5 σ. The ATP molecule has very well defined electron density. Side chains of Lys160, Ser196 and Arg198 moved relative to those in the apo-enzyme to interact with the phosphate groups, and thus have positive initial difference electron density. The figure is prepared with PYMOL [20].

Two His residues of the HIGH motif [5, 19, 21] interact with the phosphates of ATP. The first histidine, His44, has a hydrogen bond extending from its Nε2 to O1β of β-phosphate. The His47 side chain has a hydrogen bond from its Nε2 to the bridge oxygen between the α- and β-phosphate, suggesting that it functions as a general acid by donating a proton to the leaving pyrophosphate group. Both histidine residues have the Nδ1 atom forming a hydrogen bond to a main chain NH group, and thus the Nε2 atoms are hydrogen bond donors. Their imidazole rings are in close proximity to each other with the closest separation being ~3.2 Å, indicating favorable π-π interactions. These two side chains also play a role in binding the adenine group and stabilizing the reaction intermediate, because the imidazole rings are close to the adenine (closest distances of ~3.4 and ~4.3 Å for His44 and His47, respectively). Mutation of either histidine residue to alanine causes a greater than 1000-fold reduction in enzyme activity [22]. The Gly46 residue is highly conserved in this motif. It flanks one side of adenine (with Lys160 on the loop after β6 on the other side), where any side chain larger than a hydrogen atom would have steric hindrance. The KSMKS motif [23] is not conserved at the sequence level in PS enzymes, but it is evident from structural alignments [5]. Residues Ser196, Ser197, and Arg198, which are at the N-terminus of helix $3_{10}7$ and are part of the KSMKS motif, bind the β- and γ-phosphate groups. The side chains of Lys160 and Arg198 are partially disordered in the apo-enzyme structure, but become ordered and move to enable salt-bridges to the β- and γ-phosphate groups, respectively. The side chain of Lys160 also packs against one side of adenine. Mutation of this

residue to alanine reduces the enzyme activity more than 1000-fold, and it decreases affinity of the enzyme for ATP [22]. Another residue that binds ATP is Asp161, which has hydrogen bonds to two water ligands of the magnesium ion, and to O2* of the ribose group. Overall, the ATP molecule is rigidly held in the active site by many hydrophobic, π-electron, hydrogen bonding, and charge-charge interactions, which keep it in an optimal conformation for a nucleophilic attack by pantoate. Residues involved in binding the adenosine group are also important in stabilizing the pantoyl adenylate intermediate.

The binding interactions of ATP with protein are similar to those of AMPCPP [10]. However, the methylene group in AMPCPP replaces the bridge oxygen between the α- and β-phosphate groups of ATP, and thus it is not capable of forming a hydrogen bond with the Nε2 atom of His47. Consequently, the α-phosphate moves slightly away from amide nitrogen of Met40, and thus no hydrogen bond is formed between them in the AMPCPP complex. Therefore, ATP has slightly higher binding affinity to the enzyme active site than its non-reactive analog AMPCPP.

**4.5 Pantoate Binding Site**

In the $P2_1$ crystal that was soaked with pantoate and β-alanine, there is a full occupancy of pantoate in both active sites [10]. The binding site for pantoate is located in a pocket at the bottom of the active site cavity, next to the binding site of the adenosine group (**Figure 4a**). This is the same binding site proposed for the *E. coli* PS protein [5], although the pantoate molecule has a different conformation and therefore different binding interactions. The pantoate molecule is tightly bound at the bottom of the active site cavity, with side chains of two glutamine residues, Gln72 from helix $3_{10}3$ and Gln164 from helix α3, forming hydrogen bonds to its hydroxyl groups and one carboxyl oxygen. The two methyl groups face the hydrophobic groups at the bottom of the active site cavity, Pro38 and Phe157 side chains. The side chain of Met40 packs against the hydrophobic side of the pantoate molecule. Mutation of either glutamine residue to alanine greatly reduces the rate of pantoyl adenylate formation [22].

Soaking $P2_1$ crystals in solutions containing both AMPCPP and pantoate gave a full occupancy of AMPCPP, but with a glycerol in the pantoate binding site in subunit B, which has the flexible wall (residues 75-88) disordered. However, subunit A has a partial occupancy of both AMPCPP and pantoate, as indicated by their electron density and B factors. Increasing the concentration of pantoate in the soaking solution gave only a slightly higher occupancy of pantoate but a lower occupancy of AMPCPP [10]. All these observations indicate that AMPCPP and pantoate cannot coexist in the active site. This must be due to the steric and charge repulsions between the α-phosphate group and the carboxyl group of pantoate. Modeling both ATP and pantoate in one active site based on the fully occupied complexes will put one carboxyl oxygen atom of pantoate ~2.7 Å from the α-phosphorus atom of ATP [11]. This is essentially identical to what is observed in subunit A that has partial occupancy of both AMPCPP and pantoate. The carboxyl oxygen of pantoate is in a good position for an in-line nucleophilic attack on the α-phosphate. Because residues of the flexible wall (gate residues) are disordered in subunit B, the active site cavity is open and allows AMPCPP to diffuse easily into the active site. The extensive binding interactions of AMPCPP to the protein atoms make its binding thermodynamically more favorable than the binding of pantoate to the protein. Therefore, AMPCPP has a full occupancy in the active site cavity of subunit B of the complex with both AMPCPP and pantoate. On the other hand, the gate residues are less flexible in the subunit A due to crystal packing, thus making it kinetically less favorable for AMPCPP to bind, and allowing pantoate to have a partial occupancy in the active site of subunit A.

In the C2 crystal, the side chain of Gln72 on helix $3_{10}3$ swings out of the active site due to crystal packing, and the binding of pantoate is abolished. We observed only ATP in the active site with a full occupancy (see above ATP binding interactions) when we soaked the C2 crystal in a solution containing both ATP and pantoate [11]. There was no electron density for pantoate or pantoyl adenylate in the active site cavity. In the pantoate binding pocket, there was positive difference electron density, into which we modeled a glycerol molecule. This glycerol has a high temperature factor, suggesting that it has a low occupancy. Except for this Gln72 side chain, the rest of the binding site for pantoate is intact. In the $P2_1$ crystal, soaking with both ATP and pantoate results in a pantoyl adenylate intermediate in the active site cavity, and the disordered loop in subunit B becomes ordered [10] (see more details below). This indicates that loss of this glutamine side chain completely abolishes the enzyme catalytic activity for the first half of the reaction inside the crystal. The glutamine side chains are also involved in binding the reaction intermediate. Thus mutation of either residue could affect the stability of pantoyl adenylate intermediate. It is noteworthy that Asn69 was also found to be important for the enzyme activity [22]. This residue is not directly involved in binding substrates or the intermediate. However, it forms a hydrogen bond to the side chain of Gln72, and thus is important for keeping the Gln72 side chain in position. In addition, since both Asn69 and Gln72 are on

the one-turn $3_{10}3$ helix, mutation of Asn69 might also affect the stability of this $3_{10}$ helix and hence the position of the Gln72 side chain.

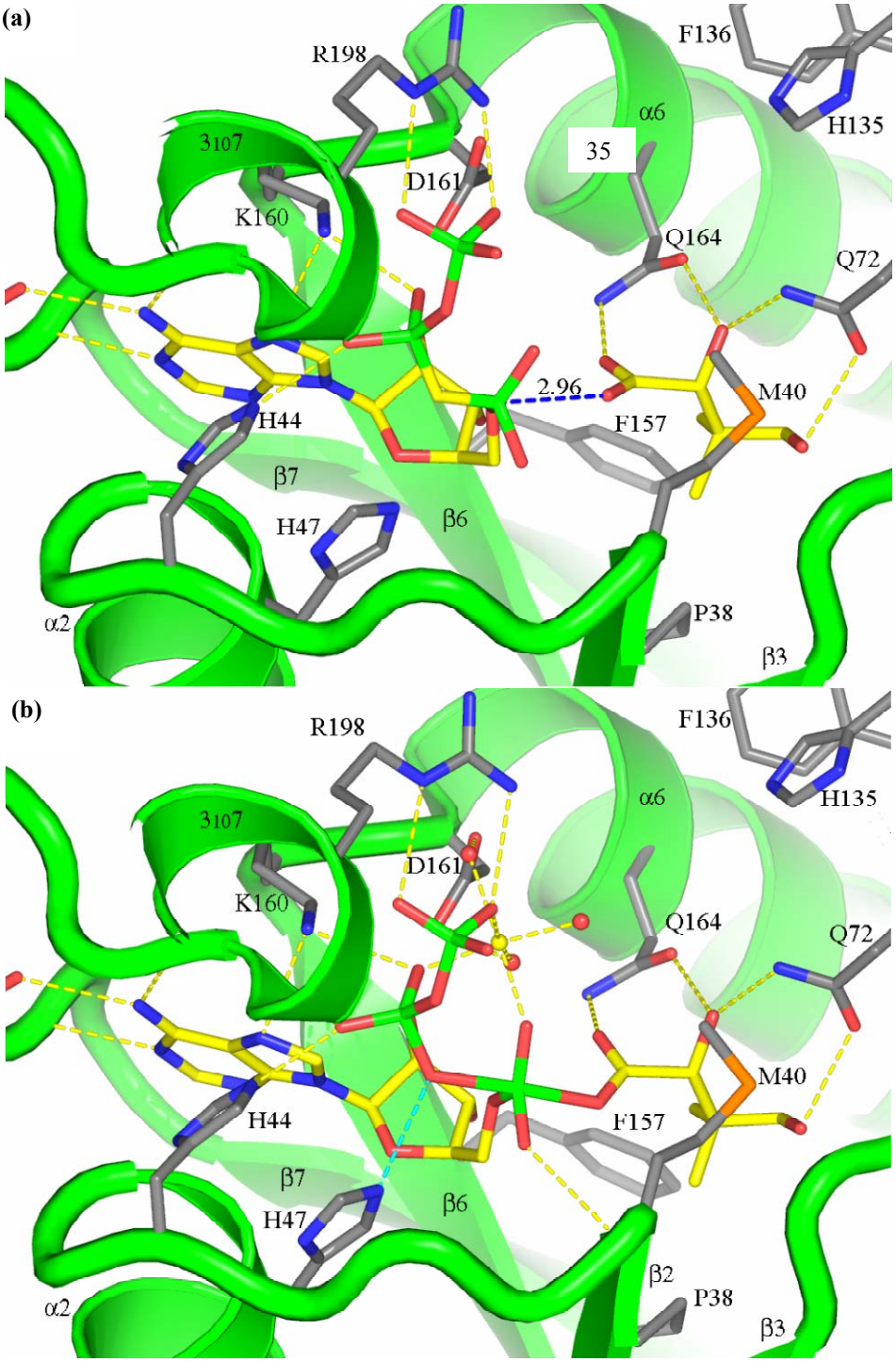



**Figure 4**. **(a)** Binding interactions of AMPCPP and pantoate in the active site of subunit A of the $P2_1$ crystal soaked in a solution containing both AMPCPP and pantoate. Both molecules have a partial occupancy (see text for details), and the magnesium ion is not well defined in this active site. However, binding position and interactions are essentially identical to those of AMPCPP and pantoate with a full occupancy when each is bound in the active site alone. The distance between the α-phosphorus atom and the carboxyl oxygen is ~2.96 Å, shown in blue dashed line. **(b)**

A model of the trigonal bipyramidal intermediate based on crystal structures of the ATP complex, AMPCPP complex and pantoate complex. Formation of the intermediate requires only the movement of the α-phosphorus atom towards the carboxyl oxygen. All favorable interactions with active residues are preserved. His47 is poised to function as a general acid.

### 4.6 Binding Interactions of the Reaction Intermediate, Pantoyl Adenylate

Soaking the $P2_1$ crystals with solutions containing both ATP and pantoate, or crystallization in the presence of both substrates, resulted in a reaction intermediate, pantoyl adenylate, in the active site [10]. Cocrystallization resulted in both active sites in the dimer being fully occupied with the pantoyl adenylate molecule. However, crystals soaked with ATP and pantoate have one active site (subunit B) fully occupied but the other partially occupied, indicated by their difference in B factors. Subunit A of the dimer also has a lower occupancy of AMPCPP when crystals were soaked in solutions containing AMPCPP. This is due to crystal packing, which makes the flexible wall of the active site cavity ($3_{10}$3-loop-α3'-loop region) ordered in subunit A and thus the active site less accessible for bulky molecules ATP and AMPCPP.

Pantoyl adenylate has extensive binding interactions with the active site residues (**Figure 5**). The molecule is almost linear and fits snugly in the bottom of the active site cavity. Both pantoate and adenylate moieties sit in the same positions in the active site as the pantoate molecule of the pantoate complex and the adenosine group of ATP in the ATP complex, and preserve all strong interactions with protein atoms. The tight binding of pantoyl adenylate stabilizes this highly reactive intermediate. One important feature of the pantoyl adenylate complex is that the disordered loop in subunit B becomes ordered, in both cocrystallization and soaking the apo-enzyme crystals with both ATP and pantoate. This suggests that the flexible wall can function as a gate, opening for substrates, especially the bulky ATP molecule, to bind and closing the active site cavity to protect the highly reactive pantoyl adenylate.

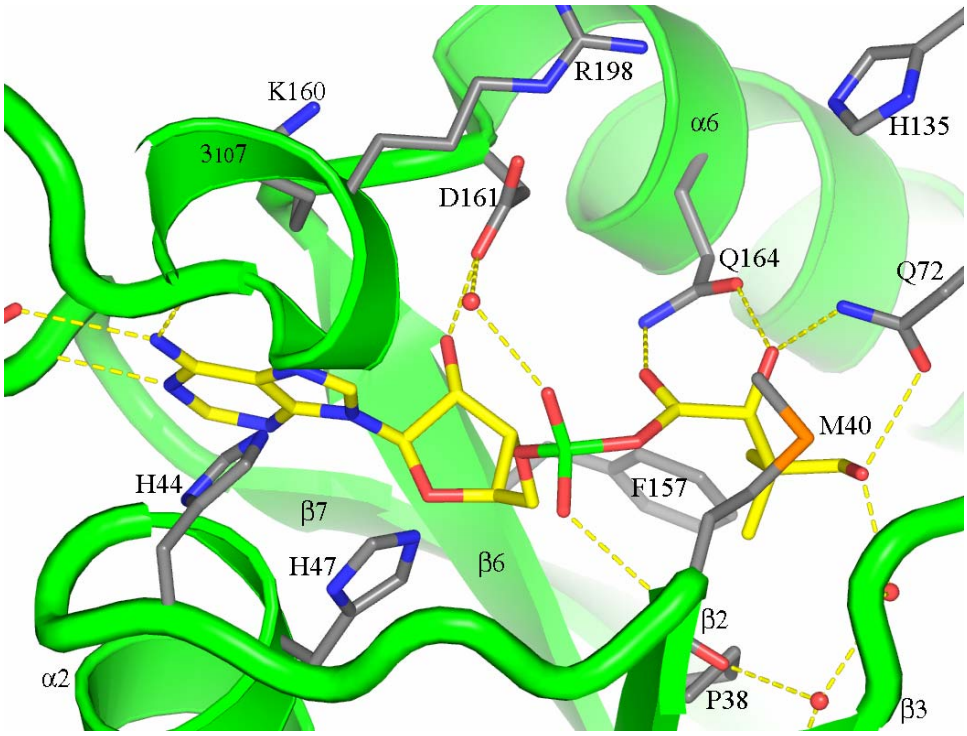



**Figure 5**. Binding interactions of the pantoyl adenylate intermediate. The intermediate binds tightly at the bottom of the active site cavity with many hydrogen bonds and hydrophobic interactions. Hydrogen bonds to the O3* of ribose and some water mediate hydrogen bonds are not shown for clarity.

### 4.7 Binding Interactions of Beta-Alanine

Cocrystallization and soaking crystals with β-alanine [10] failed to yield a complex with this substrate, suggesting that the binding site for β-alanine exists only after pantoyl adenylate is formed in the active site. Based on the structure of the pantoyl adenylate complex, it is likely that the phosphate group of pantoyl adenylate serves as an anchor for the initial binding of β-alanine by offering hydrogen-bonding partners and/or favorable charge-charge interactions. Furthermore, an AMP molecule bound at the active site may also assist the binding of β-alanine from

its phosphate group. Therefore, we soaked crystals of PS in a solution containing both β-alanine and AMP, and we obtained a crystal structure of the PS enzyme in complex with β-alanine and AMP from the $P2_1$ crystals [11].

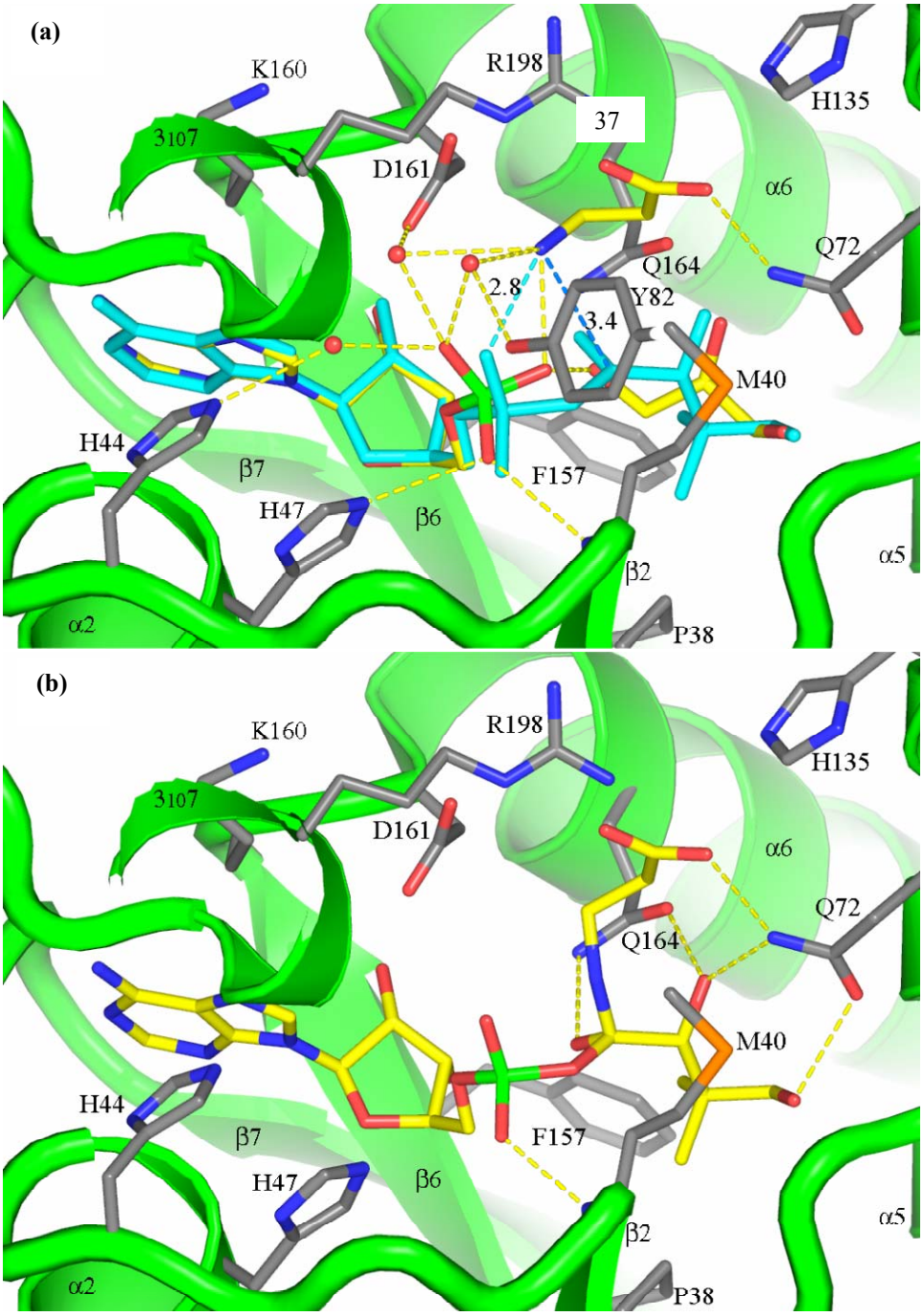



**Figure 6**. **(a)** Structural superposition of the active site of the pantoyl adenylate complex with that of the β-alanine/AMP complex. The active site residues align well with an rmsd of Cα less than 0.15 Å. The pantoyl adenylate molecule is colored in cyan. The protein structure shown in the figure is from the β-alanine/AMP complex. Hydrogen bonds to the phosphate group of AMP and to β-alanine are shown in yellow dashed lines. Some water mediated hydrogen bonds are not shown for clarity. The Tyr82 side chain is from a loop in the front that is clipped away to reveal the bound AMP and β-alanine. Based on the structural alignment, the amino group of β-alanine is ~2.8 Å from the phosphate oxygen of pantoyl adenylate, and ~3.4 Å from the carbonyl carbon, shown in blue dashed lines. **(b)** Model of the tetrahedral intermediate structure based on the structural superposition in (A). The Tyr82 side chain is not shown for clarity. A tetrahedral geometry can be obtained by a torsional rotation of the amino nitrogen of β-alanine to move it to ~1.5 Å above the carbonyl carbon. All favorable interactions with protein atoms are preserved, and thus the intermediate is stabilized by the enzyme.

The β-alanine molecule binds in the active site of subunit A in the upper part of the active site cavity, with its amino group near the phosphate group of AMP (**Figure 6a**). The amino group forms hydrogen bonds to the phosphate group of AMP and to two water molecules. Both water molecules are fixed in position through one hydrogen bond to the phosphate group and a second hydrogen bond to the side chain of Asp161 or Tyr82. The carboxyl group of β-alanine has one hydrogen bond to the Nε2 atom of th[illegible]72 side chain. This carboxyl is also close to the side chains of Arg198 and His135, with distances of ~4.0 [illegible] Å, respectively, indicating favorable charge-charge interactions and π-electron interactions. One side of β-alanine faces side chains of Met40 and Tyr82; the other side faces the large cavity of the active site, which is filled with a few ordered water molecules in the crystal structure. Binding of β-alanine is not as tight as that of other substrates and its binding site in the crystal is not fully occupied. However, the binding interactions limit molecules that can bind well and have a nucleophilic attack on the reaction intermediate. Therefore, structural analogs of β-alanine are all poor substrates [8].

In subunit B of the dimer, where residues 74 to 83 (including Tyr82) are disordered, only AMP but no β-alanine was found in the active site [11]. Also in this case, disorder of this active site cavity wall weakens binding interactions for β-alanine because its binding site is exposed to bulk solvent and thus the charge-charge interactions are dampened. Similarly, in the C2 crystal, which has one molecule per asymmetry unit and the flexible loop disordered, only AMP was found in the active site when crystals were soaked in a solution containing both AMP and β-alanine.

Superposition of the structure of the β-alanine/AMP complex with that of the pantoyl adenylate complex (**Figure 6a**) shows that the amino group of β-alanine was positioned above the carbonyl group of pantoyl adenylate such that the nitrogen atom was ~3.4 Å from the carbonyl carbon. The O1α atom of pantoyl adenylate can form a hydrogen bond with the amino group of β-alanine, serving as an anchor for the initial binding of β-alanine. With changes in only one torsion angle, the amino group of β-alanine can move towards the carbonyl carbon for a nucleophilic attack and attain a nearly perfect geometry for a tetrahedral intermediate (**Figure 6b**).

### 4.8 AMP Binding Interactions

As described above, we obtained AMP complexes from both $P2_1$ and C2 crystals by soaking crystals in a solution containing both AMP and β-alanine. Binding interactions of AMP with the active site residues in these complexes are essentially identical. The adenosine group of AMP binds in the same position as that of the pantoyl adenylate, AMPCPP, and ATP: it fits snugly in its binding pocket at the bottom of the active site cavity through hydrophobic and hydrogen bonding interactions. The adenine group is flanked by Gly46 on helix α2 and Lys160 on the loop after β6. Its N1 and N6 atoms have hydrogen bonds to main chain atoms. The hydroxyl groups of ribose form hydrogen bonds with the Asp161 side chain and a few main chain atoms at the bottom of the active site cavity. The phosphate group, however, has torsional flexibility, and it rotates slightly relative to the α-phosphate of ATP, allowing one of its oxygen atoms to form hydrogen bonds to the side chain of His47 and to the amide nitrogen of Met40 simultaneously (**Figure 6a**).

## 5. OVERALL ENZYME-CATALYZED REACTION MECHANISM

Putting all structural data together allows us to have a detailed view of the overall enzyme-catalyzed reaction mechanism. When the active site is empty, the disordered loop opens the active site and thus allows the ATP molecule to come in and bind. Because ATP is relatively bulky and is held rigidly in the active site by many binding interactions, it is likely that ATP binds first. The smaller sized pantoate then comes in, and initiates a nucleophilic attack on the α-phosphate while it binds in its binding pocket [8, 24]. When both ATP and pantoate are present in the same active site, the nucleophilic reaction must occur (**Figure 4a**). Formation of the transient trigonal bypyramidal intermediate involves only the movement of the phosphate atom towards the carboxyl oxygen of pantoate to be coplanar with the three equatorial oxygen atoms (**Figure 4b**). Although transient in nature, this intermediate is stabilized by the enzyme because all the favorable interactions are preserved, and the unfavorable steric hindrance and charge repulsions between ATP and pantoate are eliminated. The trigonal bipyramidal intermediate then dissociates, and simultaneously the pyrophosphate group leaves the active site. The magnesium ion and the positively charged side chains around the β- and γ-phosphate groups draw the negative charges towards the leaving pyrophosphate. His47 is also likely to facilitate the reaction by donating a proton to the leaving pyrophosphate. This leaves a reaction intermediate, pantoyl adenylate in the active site and completes the first half of the enzyme catalyzed reaction. Pantoyl adenylate is stabilized by many strong binding interactions, as we have seen earlier. It is also protected by the closing of the active site when the flexible loop becomes ordered. Pantoyl adenylate is stable only when tightly bound in the enzyme active site. In solution, it decomposes rapidly to give

pantoyl lactone and AMP [9]. The dissociation of the trigonal bipyramidal intermediate necessitates the simultaneous leaving of the pyrophosphate, because there would be unfavorable charge and steric repulsions once pyrophosphate is formed. This reaction, however, is reversible, as was demonstrated by positional isotope exchange experiments that $^{18}O$ labels scrambled when incubating [$\beta\gamma$-$^{18}O_6$]-ATP and pantoate with the MTB PS enzyme [24].

The presence of pantoyl adenylate intermediate and the ordering of the flexible wall create the binding site for β-alanine, the last substrate. The β-alanine molecule is relatively small, capable of passing through the small opening at the top of the active site cavity. The phosphate group of pantoyl adenylate can have a hydrogen bond to the amino group of β-alanine, thus serving as an initial anchor for binding β-alanine. The carboxylate group of β-alanine has favorable interactions with the side chains of Arg198 and His135 and a hydrogen bond to Gln72. There is a shift of the Arg198 side chain of ~1.8 Å towards the carboxylate group of β-alanine relative to that in the pantoyl adenylate complex. This binding position for β-alanine puts its amino group at a good position for a nucleophilic attack on the carbonyl carbon of the pantoyl group. The distance of the amino nitrogen is only ~3.4 Å from the carboxyl carbon of pantoyl adenylate. A model of the tetrahedral intermediate can be readily prepared from the crystal structures of the pantoyl adenylate complex and the β-alanine/AMP complex (**Figure 6b**). Formation of the tetrahedral intermediate requires only a torsional rotation of the amino nitrogen atom of β-alanine toward the carbonyl carbon of pantoyl adenylate. Similar to that of the trigonal bypyramidal intermediate, formation of the tetrahedral intermediate preserves all favorable binding interactions of active site residues with pantoyl adenylate and β-alanine. Therefore, the intermediate is stabilized by the enzyme. Dissociation of the tetrahedral intermediate then forms pantothenate and AMP. The planarity of the peptide bond of pantothenate causes significant rearrangement of the molecule, and several hydrogen bonds to active site residues are broken. In addition, the newly formed pantothenate has steric clashes and charge repulsions with AMP and some active site residues. Therefore, pantothenate must leave the active site once it is formed. The AMP molecule then diffuses away to regenerate the enzyme active site for next cycle of catalysis. Pantothenate has very low binding affinity to the enzyme. We have tried soaking and growing crystals in the presence of pantothenate, but we failed to obtain a complex with this compound. This is consistent with the finding that pantothenate is a poor inhibitor of the enzyme [8]. The AMP molecule, on the other hand, has good binding interactions in the active site. However, it typically has a low concentration inside cells, and thus can easily diffuse out of the active site. Moreover, ATP has many favorable interactions in the active site of the enzyme, and it can readily displace AMP.

## 6. REFERENCES


[1] S. Jackowski. Escherichia coli and Salmonella typhimurium: Cellular and molecular biology, American Society for Microbiology, Washington, DC 1996, pp. 687-694.

[2] W.K. Maas, Proc Int Congr Biochem 11 (1960) 161-168.

[3] V.K. Sambandamurthy, X. Wang, B. Chen, R.G. Russell, S. Derrick, F.M. Collins, S.L. Morris and W.R. Jacobs, Jr., Nat Med 8 (2002) 1171-4.

[4] W.K. Merkel and B.P. Nichols, FEMS Microbiol Lett 143 (1996) 247-52.

[5] F. von Delft, A. Lewendon, V. Dhanaraj, T.L. Blundell, C. Abell and A.G. Smith, Structure (Camb) 9 (2001) 439-50.

[6] U. Genschel, C.A. Powell, C. Abell and A.G. Smith, Biochem J 341 ( Pt 3) (1999) 669-78.

[7] A. Perez-Espinosa, T. Roldan-Arjona and M. Ruiz-Rubio, Mol Genet Genomics 265 (2001) 922-9.

[8] R. Zheng and J.S. Blanchard, Biochemistry 40 (2001) 12904-12.

[9] T. Wieland, W. Loewe, A. Kreiling and G. Pfleiderer, Biochem Z 339 (1963) 1-7.

[10] S. Wang and D. Eisenberg, Protein Sci 12 (2003) 1097-108.

[11] S. Wang and D. Eisenberg, Biochemistry (2005) In press.

[12] Z. Otwinowski and W. Minor, Methods Enzymol 276 (1996) 307-326.

[13] J. Navaza, Acta Crystallogr A 50 (1994) 157-163.

[14] A.T. Brunger, P.D. Adams, G.M. Clore, W.L. DeLano, P. Gros, R.W. Grosse-Kunstleve, J.S. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson and G.L. Warren, Acta Crystallogr D Biol Crystallogr 54 ( Pt 5) (1998) 905-21.

[15] G.N. Murshudov, A.A. Vagin, A. Lebedev, K.S. Wilson and E.J. Dodson, Acta Crystallogr D Biol Crystallogr 55 ( Pt 1) (1999) 247-55.

[16] T.A. Jones, J.Y. Zou, S.W. Cowan and Kjeldgaard, Acta Crystallogr A 47 ( Pt 2) (1991) 110-9.

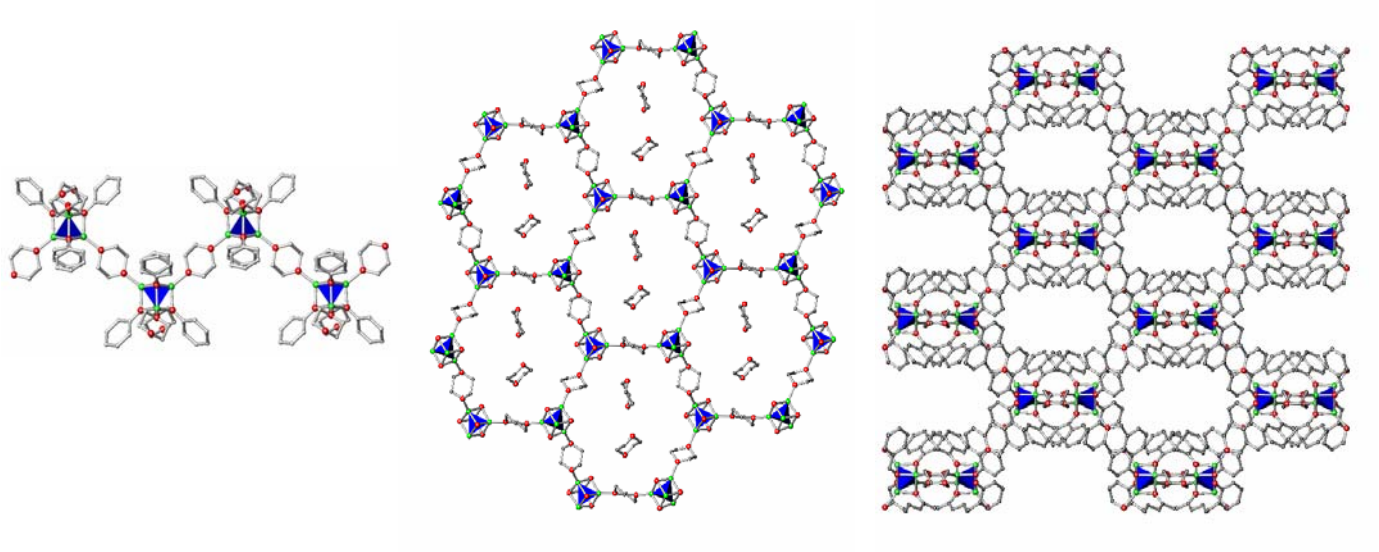
[17] P.J. Kraulis, Journal of Applied Crystallography 24 (1991) 946-950.

[18] E.A. Merritt and M.E. Murphy, Acta Crystallogr D Biol Crystallogr 50 (1994) 869-73.

[19] D.G. Barker and G. Winter, FEBS Lett 145 (1982) 191-3.
[20] W.L. DeLano, The PyMOL molecular graphics system, DeLano Scientific, San Carlos, CA, 2002.
[21] P. Bork, L. Holm, E.V. Koonin and C. Sander, Proteins 22 (1995) 259-66.
[22] R. Zheng, T.K. Dam, C.F. Brewer and J.S. Blanchard, Biochemistry 43 (2004) 7171-8.
[23] C. Hountondji, P. Dessen and S. Blanquet, Bio 40 e 68 (1986) 1071-8.
[24] L. Williams, R. Zheng, J.S. Blanchard and F.M. Raushel, Biochemistry 42 (2003) 5108-13.

*Proceedings from the Symposium*

# Margaret C. Etter Early Career Award



## ETTER TRANSACTIONS OF THE AMERICAN CRYSTALLOGRAPHIC ASSOCIATION

Edited by Frances Jurnak

VOLUME 1, 2005

*Proceedings from the Symposium*

# Margaret C. Etter Early Career Award

ACA ETTER TRANSACTIONS
VOLUME 1

Walt Disney World, Orlando, Florida
May 28 – June 2, 2005

**Editor**

Frances Jurnak
University of California, Irvine
Tel: 949-824-6623
Fax: 949-824-8540
jurnak@uci.edu

# BIS-CHELATION AND ANION EFFECTS INVOLVING A MOLECULE CONSTRUCTED IN THE ORGANIC SOLID STATE USING MOLECULAR TEMPLATES

Tamara D. Hamilton, Giannis S. Papaefstathiou and Leonard R. MacGillivray
*University of Iowa Department of Chemistry, Iowa City, IA 52242*

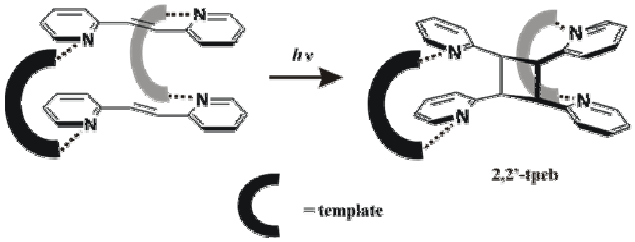
## 1. ABSTRACT

A molecule derived from a template-directed solid-state synthesis, *rctt*-tetrakis(2-pyridyl)cyclobutane (2,2'-tpcb), serves as a bis-chelating ligand upon reaction with $Cu(BF_4)_2{\cdot}H_2O$ and $NEt_4PF_6$ to produce the dinuclear complex $[Cu_2(OH)_2(H_2O)_4(\mu\text{-}2,2'\text{-tpcb})][PF_6]_2$ (**1**).

## 2. INTRODUCTION

Transition-metal-ion complexes involving seven-membered chelation rings have emerged as important for applications in catalysis and biology. Specifically, N-$C_4$-N bridged metallocycles [1] have found applications as catalysts [2] and models of biological compounds [3], as well as antibacterial [4] and antitumor [5] agents. In terms of catalysis, the application of chelating N-donor ligands - as compared to analogous polyphosphines - has begun to attract attention, especially for the activation of C-H bonds [6]. Thus, ligands that provide more than one pyridyl group in close proximity are being studied as preorganized metal chelation units [6b]. In addition, there is increasing interest in the development of novel non-metallocene catalysts for olefin polymerization for which complexes involving N-$C_4$-N bridged chelate rings have been effective [2c]. Therefore, studies that involve N-$C_4$-N ligands with a potential to offer seven-membered chelation are of much current interest.

The ability to position molecules in an orientation appropriate for reaction in the organic solid state using molecular templates that operate *via* hydrogen bonds has led to the quantitative and gram-scale construction of molecules difficult to obtain from solution [7,8]. In particular, we have shown that templates based on resorcinol can orient olefins [*e.g. trans*-1,2-bis(2-pyridyl)ethylene) (2,2'-bpe)] in the solid state in positions suitable for intermolecular [2+2] photodimerizations. The suitability of the products as ligands for coordination chemistry is evidenced by the presence of pyridine rings that emanate from central cyclobutane rings.




**Scheme 1.** Template-directed solid-state synthesis of 2,2'-tpcb from 2,2'-bpe.

As part of an ongoing study to elucidate anion effects involving transition-metal-ion complexes of *rctt*-tetrakis(2-pyridyl)cyclobutane (2,2'-tpcb), we wish to report here the synthesis and structure determination of the dinuclear Cu(II) complex $[Cu_2(OH)_2(H_2O)_4(\mu\text{-}2,2'\text{-tpcb})][PF_6]_2$ (**1**). In line with our previous study [7], the tetrapyridine acts as a bis-chelating bridge with two Cu(II) centers, giving rise to two seven-membered metallocycles. In contrast to our previous work, the $PF_6^-$ counter ions are non-coordinating and, thus, provide an opportunity for each Cu(II) ion to interact with additional donor units. In the case of **1**, each Cu(II) center is coordinated by two water molecules and a single hydroxide anion, each of which participates in hydrogen bonds with the $PF_6^-$ anion.

## 3. METHODS

Complex **1** was prepared by combining methanolic solutions of 2,2'-tpcb (0.025 g, 0.068 mmol) [8a] and $Cu(BF_4)_2{\cdot}H_2O$ (0.032 g, 0.137 mmol) (1:2 ratio). To the resulting clear blue solution was added a methanolic solution of $NEt_4PF_6$ (0.037 g, 0.137 mmol) (25 mL methanol total). A 5 mL aliquot of the latter solution was then

layered with 10 mL benzene in a test tube and allowed to sit for a period of approximately one week. The solution produced blue single crystals (yield: 0.015 g, 29 %) suitable for X-ray analysis [9].

## 4. RESULTS AND DISCUSSION

Compound **1** crystallizes in the space group $P2_1/n$. An ORTEP perspective of the asymmetric unit of **1** is shown in **Figure 1a**. The asymmetric unit consists of one Cu(II) ion, one-half molecule of 2,2'-tpcb, two water molecules, one hydroxide ion, and one $PF_6^-$. As shown in **Figure 1b**, the dinuclear complex, which sits around a crystallographic center of inversion, contains one molecule of 2,2'-tpcb that bridges two Cu(II) centers [d(Cu1···Cu1a) 7.47 Å]. The Cu(II) ions of the complex adopt an *anti* conformation and exhibit a bite angle of 88.8° [10]. Similar to $[Cu_2(NO_3)_4(\mu\text{-}2,2'\text{-tpcb})]$ and $[Cu_2(\mu_2\text{-}SO_4)_2(\mu\text{ -}2,2'\text{-tpcb})(H_2O)_2]_\infty$, each seven-membered chelation ring adopts a boat conformation. The boat conformation also compares favorably to the chelation complexes of *bis*(2-pyridyl)ethane with Pt(II) [11,12] and 1,2-bis(6-methylpyridin-2-yl)ethane with Pd(II) [13]. The coordination geometry around each Cu(II) center of **1** is square pyramidal. Two pyridyl units of 2,2'-tpcb and two water molecules (O1 and O2) form the base of the pyramid while a hydroxide ion (O3) occupies the apical position. The metal-ligand distances around the base of the pyramid range from 1.96-2.02 Å, while the Jahn-Teller distorted apical distance is 2.21 Å (Table 1).

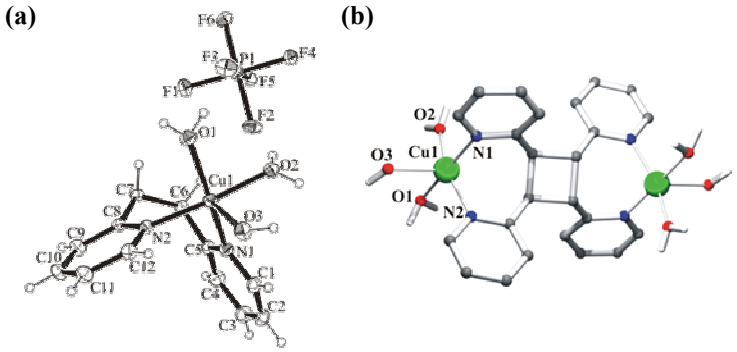



**Figure 1**. **(a)** ORTEP representation of the asymmetric unit of **1**. Displacement ellipsoids are drawn at the 30% probability level, with hydrogen atoms being assigned an arbitrary radius; **(b)** The dinuclear complex **1** (color scheme: green = copper, red = oxygen, blue = nitrogen, grey = carbon). Hydrogen atoms of 2,2'-tpcb have been omitted.

Views of the extended structure of **1** are shown in **Figures 2** and **3**. The complexes have assembled to form layers, parallel to the *ac*-plane, which exhibit an *abab* stacking pattern. The layers are held together by face-to-face π-π forces involving the ligated pyridyl groups (**Figure 2**). The counter $PF_6^-$ ions form layers sandwiched between the layers of Cu(II) complexes. The anions participate in extensive O-H···F hydrogen bonds that involve the two water molecules (O1, O2) and hydroxide ion (O3). Specifically, each $PF_6^-$ anion participates in a total of six O-H···F hydrogen bonds (**Figure 3a**) [d(O1···F3) 2.750(5) Å; d(O1···F4) 2.626(5) Å; d(O2···F2) 2.587(5) Å; d(O2···F6) 2.600(5) Å; d(O3···F1) 2.869(5) Å; d(O3···F5) 2.665(5) Å] (**Figure 3b**) with the ligands.

**TABLE 1**

Selected Bond Lengths (Å) and Angles (°) for **1**

| | | | |
|---|---|---|---|
| Cu1-O1 | 1.987(4) | O1-Cu1-O2 | 84.7(2) |
| Cu1-O2 | 1.963(4) | O1-Cu1-O3 | 99.7(2) |
| Cu1-O3 | 2.214(4) | O1-Cu1-N2 | 97.5(1) |
| Cu1-N1 | 2.020(4) | O2-Cu1-O3 | 83.7(2) |
| Cu1-N2 | 1.986(4) | O2-Cu1-N1 | 92.8(2) |
| | | O3-Cu1-N1 | 101.7(2) |
| | | O3-Cu1-N2 | 100.5(1) |
| | | N1-Cu1-N2 | 88.8(1) |

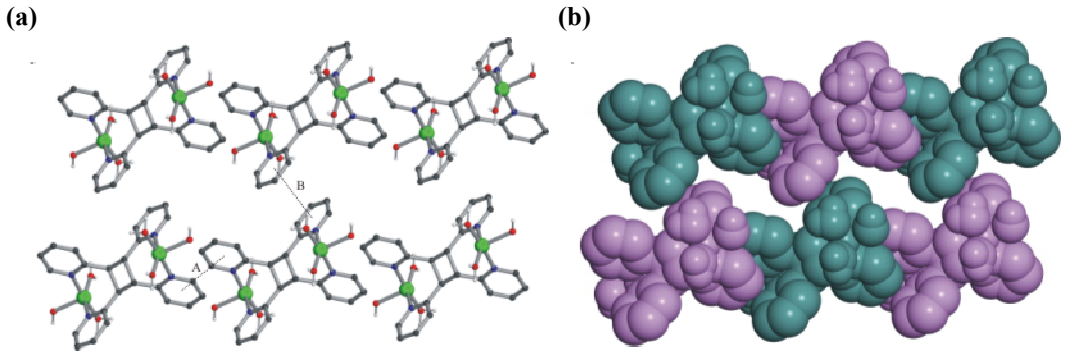



**Figure 2.** Extended structure of **1** showing the π-stacking of pyridine rings viewed in the *ac* plane: **(a)** ball-and-stick representation ($d_A$ = 3.697(5); $d_B$ = 3.874(5) Å) (color scheme: green = copper, red = oxygen, blue = nitrogen, grey = carbon) and **(b)** space-filling representation (hydrogen atoms of 2,2'-tpcb omitted for clarity).

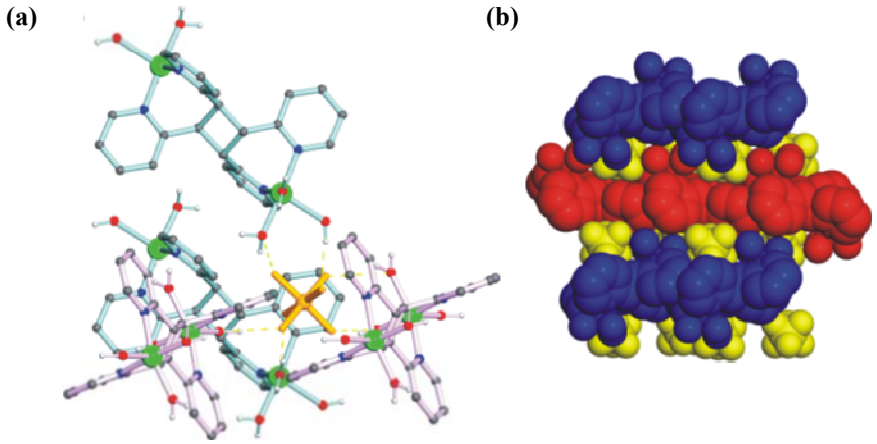



**Figure 3.** **(a)** Hydrogen bonding array involving $PF_6^-$ anion (gold) in **1**. Each F-atom is involved in one hydrogen bond (yellow) to a coordinated water or hydroxide molecule of a neighboring assembly (color scheme: green = copper, red = oxygen, blue = nitrogen, grey = carbon, white = hydrogen) (hydrogen atoms of 2,2'-tpcb omitted for clarity) and **(b)** space-filling representation of **1**, showing *abab* packing of complex layers alternating with $PF_6^-$ anions. Complexes are shown in red or blue, while $PF_6^-$ anions are shown in yellow.

## 5. CONCLUSION

In conclusion, template-directed solid-state organic synthesis has led to the construction of a tetrapyridyl ligand that forms the dinuclear coordination complex **1** with Cu(II) ions. The complex incorporates two seven-membered chelation rings [7]. 2,2'-tpcb is the first ligand to sustain seven-membered chelation in a series of coordination complexes. The non-coordinating behavior of the $PF_6^-$ counter-ion provides open sites on each metal, which are occupied by water molecules and hydroxide ions.

## 6. REFERENCES

[1] a) McCullough, F. Jr.; Bailar, J. C. Jr. *J. Am. Chem. Soc.* **1956**, *78*, 714-716;
b) Habu, T.; Bailar, J. C. Jr. *J. Am. Chem. Soc.* **1966**, *88*, 1128-1130.

[2] a) Gibson, V. C.; Kimberley, B. S.; White, A. J. P.; Williams, D. J.; Howard, P. *Chem. Commun.* **1998**, 313-314;
b) Kretzschmar, E. A.; Kipke, J.; Sundermeyer, J. *Chem. Commun.* **1999**, 2381-2382;
c) Siemeling, U.; Kolling, L.; Stammler, A.; Stammler, H. G.; Kaminski, E.; Fink, G., *Chem. Commun.* **2000**, 1177-1178;
d) Bacchi, A.; Pelizzi, G.; Minardi, G.; Pistuddi, A. M.; Solinas, C.; Chelucci, G. *Trans. Met. Chem.* **2002**, *27*, 274-278;
e) Minardi, G.; Mura, E.; Pistuddi, A. M.; Solinas, C.; Bacchi, A.; Pelizzi, C.; Pelizzi, G.; Chelucci, G *Trans. Met. Chem.* **1999**, *24*, 481-485;

f) Zhu, H. B.; Dai, Z.-Y.; Huang, W.; Cui, K.; Gou, S.-H.; Zhu, C.-J. *Polyhedron* **2004**, *23*, 1131-1137;
g) Bandoli, G.; Gerber, T. I. A.; Jacobs, R.; du Preez, J. G. H. *Inorg. Chem.*, **1994**, *33*, 178-179.

[3] a) Martin, J. W. L.; Timmons, J. H.; Martell, A. E.; Rudolf, P.; Clearfield, A. *Inorg. Chem.* **1981**, *20*, 814-821;
b) Choo P. L.; Mulichak, A. M.; Jones Jr., R. W.; Bacon, J. W.; Pett, V. B.; Zacharias, D. E. *Inorg. Chim. Acta* **1990**, *171*, 183-192;
c) Ball, R. G.; Trotter, J. *Can. J. Chem.* **1979**, *57*, 1368-1371;
d) McFarland, C. A.; Gross, S. A.; Winfield, J. S.; Pett, V. B. *Inorg. Chim. Acta*, **1994**, *221*, 35-41;
e) Balamurugan, R.; Palaniandavar, M.; Gopalan, R. S. *Inorg. Chem.* **2001**, *40*, 2246-2255.

[4] Nandi, A. K.; Chaudhuri, S.; Mazumdar, S. K. Ghosh, S. *Inorg. Chim. Acta* **1984** *92*, 235-240.

[5] a) Khokar, A. R.; Shamsuddin, S.; Xu, Q. *Inorg. Chim. Acta* **1994**, *219*, 193-197;
b) Shamsuddin, S.; van Haal, J. W.; Stark, J. L.; Whitmore, K. H.; Khokhar, A. R. *Inorg. Chem.* **1997**, *36*, 5969-5971;
c) Shamsuddin, S.; Santillan, C. C.; Stark, J. L.; Whitmire, K. H.; Siddik, Z. H.; Khokhar, A. R. *J. Inorg. Biochem.* **1998**, *71*, 29-35;
d) Hoeschele, J. D.; Showalter, H. D. H.; Kraker, A. J.; Elliott, W. L.; Roberts, B. J.; Kampf, J. W. *J. Med. Chem.* **1994**, *37*, 2630-2636.

[6] a) Vedernikov, A. N.; Huffman, J. C.; Caulton, K. G. *Inorg. Chem.* **2002**, *41*, 6244-6248;
b) Vedernikov, A. N.; Huffman, J. C.; Caulton, K. G. *Inorg. Chem.* **2002**, *41*, 6867-6874; Vedernikov, A. N.; Pink, M.; Caulton, K. G. *Inorg. Chem.* **2004**, *43*, 3642-3646.

[7] Hamilton, T. D.; Papaefstathiou, G. S.; MacGillivray, L. R. *Cryst. Eng. Comm.* **2002**, *4*, 223-226.

[8] a) Papaefstathiou, G. S.; Kipp, A. J.; MacGillivray, L. R. *Chem. Commun.* **2001**, 2462-2463;
b) MacGillivray, L. R.; Reid, J. L.; Ripmeester, J. A. *J. Am. Chem. Soc.* **2000**, *122*, 7817-7818;
c) Hamilton, T. D.; Papaefstathiou, G. S.; MacGillivray, L. R. *J. Am. Chem. Soc.* **2002**, *124*, 11606-11607.;
d) Varshney, D. B.; Papaefstathiou, G. S.; MacGillivray, L. R. *Chem. Commun.* **2002**, 1964-1965;
e) Gao, X.; Friščić, T.; MacGillivray, L. R. *Angew. Chem. Int. Ed.* **2004**, *43*, 232-236.

[9] Crystal data for **1**: monoclinic, $P2_1/n$, Z = 4, a = 9.716(5), b = 16.586(5), c = 9.939(5) Å, $\beta$ = 98.270(5) º, V = 1585.0(12) Å$^3$, 2738 unique reflections, $R$ = 0.0492 [for 2122 reflections with $I \geq 2\sigma(I)$], $R$ = 0.0742 and w$R^2$ = 0.1408 (all data). Crystallographic Data has been deposited with the Cambridge Crystallographic Data Centre and has been allocated with deposition number CCDC 291578.

[10] The bite angle is the angle between chelating ligand donor atoms (L-M-L) in a metal complex.

[11] McFarlane, A.; Lusty, J. R.; Fiol, J. J.; Terron, A.; Molins, E.; Miravitlles, C.; Moreno, V. *Z. Naturforsch., Teil B*, **1994**, *49*, 844-848.

[12] a) Marcelis, A. T. M.; Korte, H.-J.; Krebs, B.; Reedijk, J. *Inorg. Chem.*, **1982**, *21*, 4059-4063;
b) Kleibohmer, W.; Krebs, B.; Mercelis, A. T. M.; Reedijk, J.; Van der Veer, J. L. *Inorg. Chim. Acta* **1983**, *75*, 45-50.

[13] Baker, A. T.; Crass, J. K.; Kok, G. B.; Orbell, J. D.; Yuriev, E. *Inorg. Chim. Acta* **1993**, *214*, 169-176.

# USE OF ALKALI METAL AGGREGATES IN CONTROLLING NETWORK ASSEMBLY

Dugald J. MacDougall, J. Jacob Morris, Bruce C. Noll and Kenneth W. Henderson*
*Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, Indiana 46556, USA.*

## 1. ABSTRACT

This paper summarizes the first systematic studies into the use of *s*-block aggregates to control the rational assembly of network polymers. Preliminary work using lithium and sodium aryloxide aggregates has established the efficacy of this methodology, and in particular that the control of network architectures can be achieved on a regular basis. These studies also demonstrate that air-stable, robust, open-framework materials may be prepared through use of these systems. Overall, the paper outlines an area that has considerable potential for the preparation of new materials, that advance knowledge in the understanding of the self-assembling process.

## 2. INTRODUCTION

The synthesis and characterization of periodic network assemblies is an area of increasing international academic and industrial interest [1-4]. The attention paid to this topic can be attributed, in part, to the fundamental scientific challenges present in developing rational routes to structurally well-defined solids and also as a consequence of the potential utility of these materials in applications as diverse as catalysis [5], chemical separation [6], optics [7] and electronics [8]. In this regard, two main strategies have been adopted in the formation of frameworks from molecular precursors [9], firstly through the use of hydrogen-bonded organic solids [10] and secondly using metal-organic coordination polymers containing transition elements and ligand spacer molecules [11]. These strategies have also been combined where metal-containing complexes are incorporated within hydrogen-bonded networks [12]. In addition, a notable area of success in controlling the formation of covalently-linked solids has been the use of metal carboxylate clusters as directional components of metal-organic frameworks (MOFs) [13]. However, a prominent omission from the current list of strategies directed towards rational network synthesis is the use of early main group metals to dictate supramolecular structure. This deficit may be explained by the considerable potential problems associated with these species, such as the presence of complex dynamic solution equilibria between aggregated complexes; the spherical non-polarizable nature of the cations; the limited preference for fixed metal geometries, and the lability of the metal-ligand bonding. All of these issues may lead to poor predictability and reproducibility in network synthesis [14-16]. In this context it is unsurprising that researchers have opted for alternative structural scaffolds, and to a large extent have consciously avoided the use of *s*-block metal cations. In this paper we directly address these established perceptions and contest that the early main group elements can be successfully employed to control network assembly by the informed selection of metal, anionic partner, linker molecule and synthetic conditions. More specifically, we outline a novel approach to rational network synthesis utilizing pre-assembled *s*-block molecular aggregates as secondary building units (SBUs) to control supramolecular structure. Although the use of SBUs is quickly emerging as a successful tactic in rationalizing and controlling the formation of various types of extended framework architectures, the use of *s*-block SBUs has not been investigated [17]. In this paper we detail our preliminary investigations in this area, which firmly establish the efficacy of this methodology. We will outline an unprecedented systematic series of one-, two- and three-dimensional polymers constructed from *s*-block metal aggregates. A subset of these networks displays intriguing host-guest chemistry, including the preparation of air-stable, robust, open-framework materials. We have also developed a model for network prediction, which will be an essential element in our future studies this area.

### 2.1 Initial Observations and Proof-of-Concept Preliminary Studies

During the course of our investigations into the structure and function of $\alpha$, $\alpha$'-stabilized carbanions we discovered that the lithium salts of many of these species form unusual polymeric materials [18]. We structurally characterized lithiated $\alpha$-cyanophosphonates of the type $[(RO)_2P(O)CHCNLi.THF]$ (R = Et or $Pr^i$) and found that they form 2D sheet structures, as displayed in **Figure 1**.

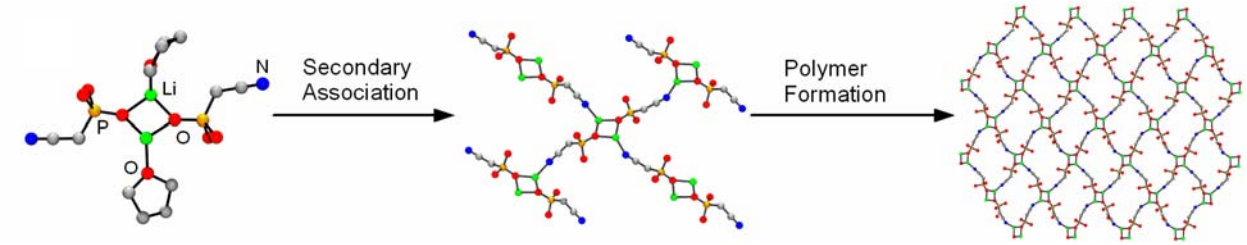



**Figure 1.** Assembly process of lithiated α-cyanophosphonates by linking $Li_2O_2$ dimers.

Intriguingly from the molecular perspective, the sheets can be described as being composed of 'dimeric' $Li_2O_2$ units that are interconnected to four identical 'dimers' via ligation of the nitrile groups. The $Li_2O_2$ motif, where the metals bridge between a pair of phosphoryl units, is entirely consistent with established aggregation patterns for simple lithiated phosphonates [19]. Then, the pendant nitriles act as linear-linkers and bond in an 'interdimer' fashion to another $Li_2O_2$ unit, with the adjacent rings rotated by approximately 90° with respect to one another. Overall, four dimers interconnect through the nitrile bridges to give large $(LiOPCCN)_4$ twenty-four membered rings with small four membered $Li_2O_2$ rings at the corners. This pattern is topologically equivalent to a (4,4) net and can be described as a decorated network where one square-planar vertex has been replaced by a molecular aggregate [20, 21]. The networks obtained are reminiscent of classical structures utilizing square-planar transition metal nodes, such as $Ni^{2+}$, $Zn^{2+}$, $Cu^{2+}$, $Cd^{2+}$, *etc.*, with neutral linear-linker ligands such as 4,4'-bipyridine. In our case the $Li_2O_2$ dimeric aggregates act as square-planar nodes, where the nodal point is located at the center of the ring rather than being an individual atom.

We theorized that it should be possible to form related architectures in a rational and controlled manner using other types of lithiated aggregates. We chose to examine (organo)sulfonylacetonitriles, $RSO_2CH_2CN$, as carbanion sources since simple lithiated sulfonyls are known to form $(SO_2Li)_2$ eight membered ring dimers [22]. Therefore, these complexes should assemble in a similar manner to the cyanophosphonates described above but with larger eight membered rings interconnecting the macromolecular twenty-four membered rings. In this vein we targeted and subsequently successfully prepared and structurally characterized the complex $[MeSO_2CHCNLi.THF]$, [23] and we were pleased to confirm that this complex forms a network with exactly the connectivity that was predicted in advance [24, 25]. **Figure 2** illustrates the formation of this two dimensional sheet that may be described as a basket weave network. Overall, the $(SO_2Li)_2$ eight-membered ring acts as an offset square planar node to give the desired topology.

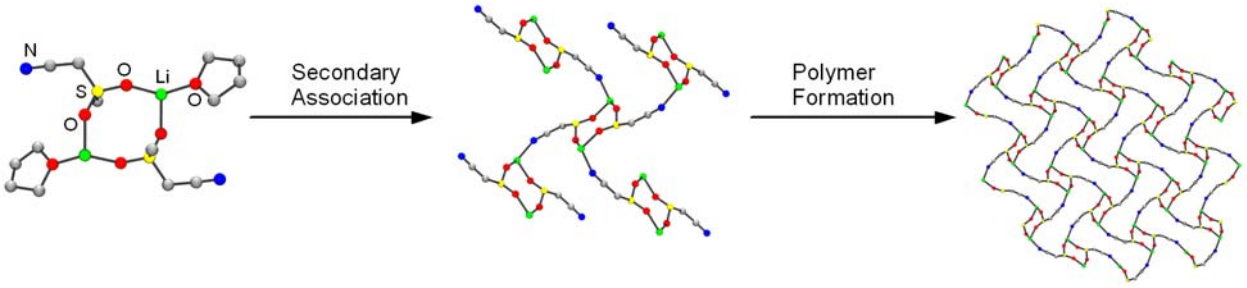



**Figure 2.** Assembly process of lithiated (organo)sulfonylacetonitriles through association of $(SO_2Li)_2$ dimers.

## 3. METHODS

The coordination chemistry involving the early main group elements is an active area of academic study, with the molecular aggregation behavior of numerous classes of metal-ligand combinations now well established [14-16]. With this understanding of molecular aggregation in hand, we aimed to select a series of readily prepared and well-behaved main group aggregates that could then be utilized as SBUs. A highly attractive feature of *s*-block SBUs is the diversity of metal aggregate geometries and compositions available as synthons. Our goal was to utilize the

geometric and chemical information stored in a variety of metal-containing ring and cage compounds to rationally design, and subsequently build, novel two- and three-dimensional network architectures. Specifically, since the location of the metal centers and the anions within these aggregates are known in advance, we could use this information to predict the directions for polymer extension. Suitable aggregates were selected that contain metals with one or more free coordination sites or labile ligands for Lewis base ligation and subsequent polymer growth. In essence the entire aggregate is used as a fixed nodal point (as opposed to the nodes being single spherical metal cations) with their geometries determined by the nature of the rings or cages. This strategy includes utilizing strong 'primary' bonding interactions to assemble molecular aggregates of predictable composition, shape and size, which are then linked via weaker 'secondary' Lewis base-type bonding. The interaggregate association is achieved by the use of neutral polydentate Lewis bases containing at least two donor sites, as opposed to the use of internal donor groups attached to the anionic ligand backbones as described in **Figures 1** and **2**. The use of SBUs in network synthesis is also an appealing approach since they dramatically reduce the number of possible network topologies arising for a given node/linker combination and they may also aid in the preparation of stable porous materials [26].

It is pertinent to recognize that polymer formation is in fact very common for early main group compounds, in particular for *s*-block complexes [27, 28]. Indeed, most preparative chemists have from time to time encountered an insoluble lithiated intermediate during a synthesis. In many instances such solubility problems can be overcome by the use of polar solvents such as THF, DME and TMEDA, which cleave metal-ligand interactions and generally reduce the aggregation state of the metallated species. In turn, the vast majority of work carried out on the structure and bonding of early main group metal complexes has been performed at the molecular level, and the structural elucidation of polymeric materials has mainly been by inference rather than by direct analysis [29]. Although, various types of coordination polymers containing *s*-block metals have previously been characterized, they have usually been prepared either inadvertently or in order to study localized metrical information, rather than deliberately to build specific network architectures [30]. Indeed, while elegant work has been carried out elucidating the subtle factors involved in determining the molecular structures adopted by *s*-block metal complexes, no sustained attempt has been made to rationalize polymer formation in this area.

### 3.1 Common Geometries Available for Alkali Metal Containing SBUs

The use of lithiated SBUs was a reasonable starting point for the study since the understanding of the aggregation behavior for these complexes is by far the most developed within the *s*-block series [14-16]. Our initial phase of the project utilized the abundance of information that is available for lithiated species to choose suitably robust molecular aggregates as building blocks. **Figure 3** displays the core structures of a selection of commonly found aggregation states for *s*-block complexes: ring dimers, ring trimers, tetrameric cubanes, and prismatic hexamers. **Figure 3** also illustrates the relationship between the metal arrangements within the aggregates and their corresponding nodal geometry. The exact arrangement of the metals within each type of aggregate will vary with the identity of its anionic partner. This will lead to a variety of geometric distortions from the 'ideal' nodal configuration. Nevertheless, the basic ring or cage arrangements should remain intact and will therefore control the direction of polymer assembly.

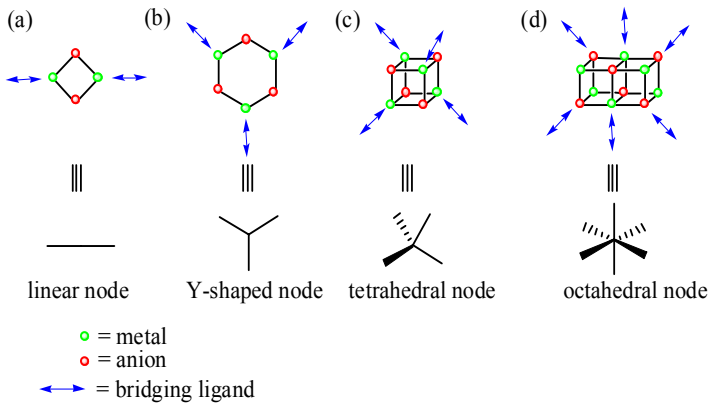



**Figure 3.** Relationship between selected metal aggregates and their related nodal geometries: (a) disolvated ring dimer, (b) trisolvated ring trimer, (c) tetrasolvated tetrameric cubane, and (d) hexasolvated prismatic hexamer.

### 3.2 Lithium Aryloxides as SBUs

It was essential to the success of our approach that the aggregation state of the metallated compound was predictable and remains intact both in solution and in the solid-state while in the presence of the Lewis base linker. An excellent starting point was lithium aryloxides. These complexes contain strong Li-O bonding, and numerous aggregate types are available as SBUs. In addition, they are readily prepared from the parent phenol by direct deprotonation using bases such as organolithium or lithium amide reagents. The aggregate size and coordination environment around the metal center can readily be controlled by the choice of phenol and donor solvent. For example, increasing the steric bulk at the 2,6-positions of an aryloxy ring results in decreasing the aggregation state of the lithiated complex. This is conveniently demonstrated for a series of THF-solvated lithiated aryloxides (ArOLi), shown in **Figure 4**, which form stable ring dimers (Ar = 2,6-$^{t}Bu_2C_6H_3$), ring trimers (Ar = 2,6-$^{i}Pr_2C_6H_3$), tetrameric cubanes (Ar = 2,4,6-$Me_3C_6H_2$), and prismatic hexamers (Ar = $C_6H_5$) [31-33].

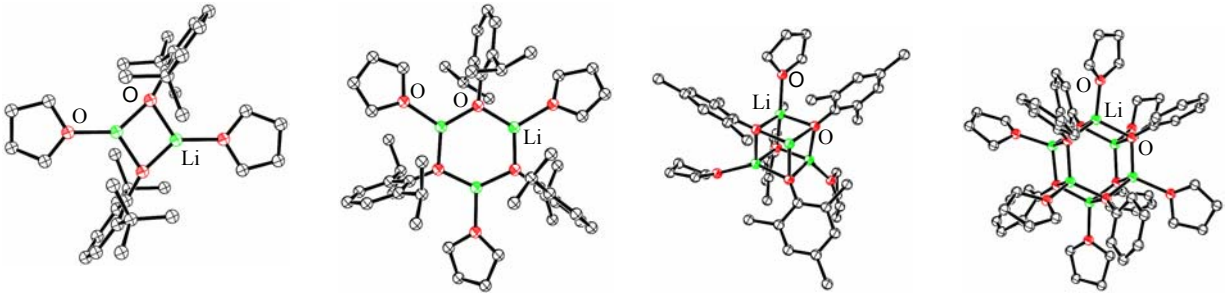



**Figure 4.** Molecular structures of a series of THF-solvated lithium aryloxides, $[ArOLi.THF]_n$.

Each of the metal atoms within these aggregates has one coordination site occupied by a monodentate THF donor molecule. This makes these anionic aryloxide ligands excellent targets as SBUs since replacement of the terminal Lewis bases by didentate linear-linker molecules will result in interconnecting aggregates. There are of course numerous possibilities for both the connectivity and topology of the supramolecular assemblies that may be formed via the interaction of the four aggregate types shown above with a linear-linker such as dioxane. However, it is becoming clear that a relatively few, high-symmetry structures dominate network topologies [26]. The most likely assemblies to be produced in each case are one-dimensional chains from disolvated dimers; hexagonal (6,3) sheets from ring trimers; three-dimensional diamondoid networks using tetrameric cubanes; and cubic lattices using prismatic hexamers.

### 3.3 Choice of Neutral Linker Ligand

Polydentate Lewis bases are required in order to connect neighboring aggregates. The strategy used to achieve this goal was to employ neutral linear-linker ligands containing relatively hard Lewis base donors [34]. A selection of potentially useful linkers for the *s*-block systems are shown in **Figure 5**.

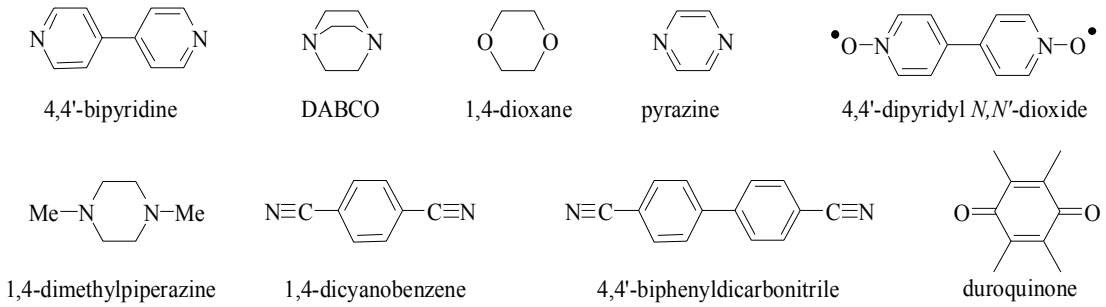



**Figure 5.** Neutral linear-linker molecules suitable for coordination to hard *s*-block metals.

Generally, divergent ligands have been used in lithiated systems simply to enhance the crystallization properties of a metallated molecular substrate in order to study localized bonding or metrical information, rather than to

deliberately build specific network structures [30, 35, 36]. However, in most instances no recognizable lithiated aggregate is present, rather isolated ions with no inherent directional control are linked together. Furthermore, reports on the use of other linear-linking didentate Lewis bases for the *s*-block metals are surprisingly scarce. This deficiency is remarkable considering the enormous importance of such ligands in the assembly of transition metal-containing networks [37]. Nevertheless, ligands of this type should be ideal candidates for coordination to *s*-block metals, in particular lithium, due to their excellent properties as hard Lewis bases. The monodentate base pyridine is a commonly used donor solvent for lithiated complexes with dozens of structurally characterized examples in the Cambridge Structural Database [38], and furthermore several crystal structures of lithium species have been characterized containing the chelating ligand 2,2'-bipyridine [39]. In the present study the linker of choice is 1,4-dioxane as this ligand proved to be convenient as it not only acts as a divergent linker but also as solvent media.

## 4. RESULTS AND DISCUSSION

### 4.1 Use of Tetrameric $Li_4O_4$ Cubanes as SBUs Linked by Dioxane

This early success in rationally designing complex network architectures from very simple constituents spurred us to expand the scope of our studies to include the possibility of linking together neutral, pre-assembled lithiated SBUs using *external* neutral divergent Lewis bases. This is appealing since it immediately opens up a wide variety of well-studied solvated aggregate types for investigation. We targeted the commonly encountered $Li_4O_4$ tetrameric cubanes formed by lithium aryloxides (ArOLi) for use as SBUs [31]. These complexes are excellent SBU candidates since they contain strong Li-O bonding, the metals are held in an approximately tetrahedral arrangement with one coordination site available for ligation, and there is a wide range of substituted phenols available for systematic studies. The didentate donor dioxane was selected as the divergent external linking Lewis base due to its capacity to act as solvent media, its rigidity and its inability to chelate a lithium center.

At the outset, an exploratory computational study was conducted to determine the feasibility of dioxane acting as a bridging ligand between a pair of $Li_4O_4$ cubanes, since this structural pattern had not previously been characterized (**Figure 6**). Semi-empirical PM3 calculations were the first line of analysis due to their computational efficiency [40, 41]. Geometry optimization calculations were performed on the full molecules [(PhOLi.dioxane)$_4$] and [{(PhOLi)$_4$}$_2$.(dioxane)$_7$] rather than using simplified model complexes to ensure an accurate portrayal of subtle steric and electronic effects. These calculations indicated that no notable energy penalty (<0.2 kcal/mol) is incurred on bridging a pair of tetrasolvated cubanes by the donor. This key result was confirmed by further geometry optimization at the *ab initio* HF/6-31G* level of theory (<0.01 kcal/mol) and also by high level single-point density functional theory calculations (B3LYP/6-311G**, <0.04 kcal/mol) [41-44]. These calculations firmly established the viability of polymer formation. In addition, this study demonstrates the ability of lower level calculations to be used as an efficient method for screening suitable aggregate/linker combinations.

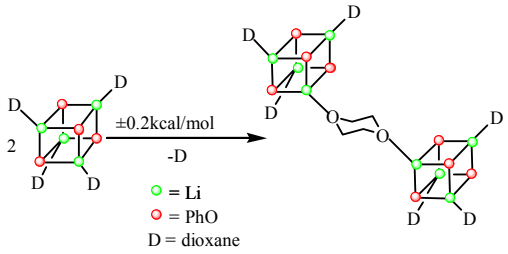



**Figure 6.** Calculated energetics of dioxane bridging a pair of $Li_4O_4$ units.

Appropriate lithium aryloxides were then chosen based on their predilection to form tetrasolvated $Li_4O_4$ cubanes in the presence of monodentate Lewis bases [23]. The required tetrameric aggregation was confirmed by preparing and structurally characterizing (by X-ray diffraction, XRD) the molecular analogues using monodentate donors (THF or pyridine), and also by solution NMR studies of the complexes in dioxane solvent media using the Jackman method of correlating the $^{13}C$ NMR shift positions for the *para*-carbons with aggregation state [25, 45]. Specific structural variants of these ligands were then selected as our understanding of these systems evolved. **Figure 7** highlights the structural characterization, by single-crystal XRD, of the three representative complexes [{(ROLi)$_4$.(dioxane)$_x$}$_\infty$], (a) R=Ph, x=3; (b) R=4-Et-$C_6H_4$, x=2.5; and (c) R=1-naphth, x=2, which typify our findings [46].

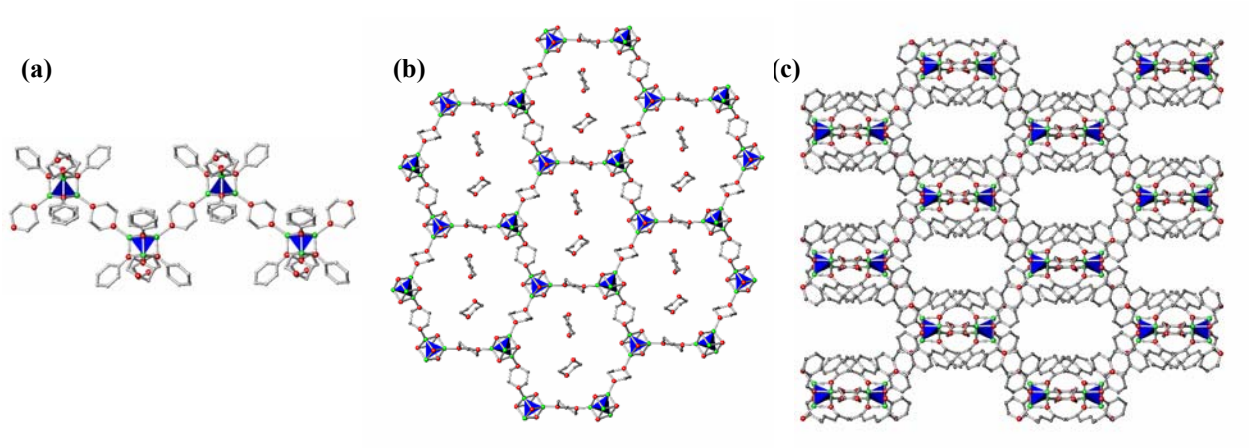



**Figure 7.** Sections of the polymeric structures of the: **(a)** 1D zig-zag chain, **(b)** framework atoms of the 2D hexagonal network with two guest solvent molecules per macrocycle, and **(c)** 3D diamondoid network. Vertices of the blue tetrahedra represent the four lithium centers within each cubane.

The first key point is that each structure is composed of linked $Li_4O_4$ tetrameric cubane units, proving that the desired molecular SBU remains intact on polymerization. Next, the three polymeric types found, 1D zig-zag chains, 2D hexagonal nets, and 3D diamondoid networks are very closely related, simply differing by the number of points of extension from the tetrahedral SBU (two, three and four leading to V-shaped, trigonal and tetrahedral nodes respectively). Moreover we have now fully characterized an extended series of $Li_4O_4$ substituted aryloxide complexes and found that they each form one of these three structural types. Pleasingly, this demonstrates that the SBU approach is indeed successful in limiting the number of network topologies for a given node/linker combination.

The complexes in **Figure 7** illustrate the dramatic effect of relatively small changes in sterics on the macromolecular structure adopted. We deduced that the type of polymer formed can be rationalized in terms of balancing entropic factors with the efficient filling of space. Although our calculations indicate that bridging by dioxane is thermoneutral, there will be a substantial increase in entropy associated with polymer formation due to the gain in translation freedom of the liberated solvent molecules. Next, there will clearly be an optimum length for the rigid anionic ligands of the SBUs in order to efficiently fill the hexameric macrocycles of the 2D sheets. If the ligands are too short this will result in energetically unfavorable large voids at the center of the rings. Conversely, if the ligands are too long the 2D arrangement will be destabilized by transannular interactions [47]. This analysis is consistent with the 1D chain structure observed for the small $PhO^-$ anions. Extension of the ligand at the *para*-position partially circumvents this problem and the longer 4-Et-$C_6H_4O^-$ anions, in combination with the guest dioxanes, fill the hexameric macrocycles with reasonable efficiency. However, examination of the extended crystal-packing diagram of this complex shows interdigitation between neighboring 2D layers, suggesting that significant space still remains within the sheet. In order for a diamondoid structure to be formed it is necessary to occupy 3D adamantanoid cavities. This proves possible for 1-naphthOLi by lateral extension to the aromatic ring, leading to excellent localized space filling in the 'corners' of the adamantanoid cavities by interlocking of the naphthyl units [47]. Space filling by interpenetration is precluded due to the $(1\text{-naphthOLi})_4$ cubanes being ~14.2 Å in diameter whereas the adamantanoid cavities have dimensions of only 7.3 x 7.6 x 16.6 Å. The open structure found for this complex supports the conjecture that the use of SBUs promotes the formation of open-framework materials. The viability and scope of our rationalization was tested through targeting the structural characterization of specific substituted aryloxides. This proved successful, such that when the *para*-substituent is small (H, Me, F or Cl) 1D chains are produced, minimal extension at this position (Et, OMe, $^i$Pr) leads to the formation of 2D hexagonal sheets, whereas lateral substitution of the aromatic (Ar = 1-naphth or 2,4,6-$Me_3C_6H_2$) gives 3D diamondoid networks. In all cases the homogeneity of the samples was confirmed by powder XRD of the bulk solids.

The potential for porosity is an exciting feature of the diamondoid structures outlined above and our preliminary investigations indicate that the guest molecules within $[\{(1\text{-naphthOLi})_4.(\text{dioxane})_2\}_\infty]$ (**Figure 7c**) can be removed under reduced pressure without disruption to the framework. Solvent loss was monitored by $^1$H NMR spectroscopy through complete dissolution of samples in $d_6$-DMSO, followed by integration of the anion and dioxane signals.

Initially, three enclatherated solvent molecules per $Li_4O_4$ unit are present but after 24 hours of evacuation ($1.5 \times 10^{-3}$ Torr and 30°C) the non-framework dioxane is entirely removed. Calculations indicate that 34.8% of the total volume within this structure is potential solvent space.[77] Comparison of the calculated, as prepared, and evacuated powder XRD patterns show only minor variations, demonstrating that the integrity of the framework is maintained on solvent loss. In addition, the macroscopic morphology of the crystals is retained after evacuation (perfect octahedra) allowing determination of the unit cell parameters by single-crystal XRD, and pleasingly they match those of the original material (a full structure determination has not yet been completed due to the weak diffraction of the crystals). Furthermore, these crystals were found to be completely air-stable over several weeks. Unit cell checks confirmed that the crystallinity is retained and IR spectroscopy showed no detectable decomposition to lithium hydroxide.

### 4.2 Use of Hexameric $Na_6O_6$ Prismatic Cages as SBUs Linked by Dioxane

Many complexes of the heavier Group 1 metals are appealing as SBUs. Of the heavier alkali metals, the alkoxides and aryloxides of both sodium and potassium have been reasonably well studied [48, 49]. A potential drawback of moving from lithium to the heavier Group 1 metals is a significant weakening of the metal-ligand bond strength, which may in turn lead to increased aggregate lability and a decrease in the predictability of the nature of the SBU. To test the feasibility of using these metal aggregates we targeted solvated $Na_6O_6$ hexameric aggregates as potential sodium based SBUs since this structural type has been characterized in the solid state for molecular derivatives, *e.g.* $[(PhONa.THF)_6]$ [50]. We theorized that the triple stack of dimers (face-shared cubane) aggregates should act as octahedral SBUs and lead to the formation of 3D cubic networks. Building on our work on the lithiated systems, we utilized 4-substituted aryloxides as convenient probes to test the effect of sterics on the supramolecular arrangements obtained. The set of *para*-halide substituted derivatives are particularly useful in illustrating the effect of sterics on network structure. The complexes $[(4\text{-R-}C_6H_4ONa)_6.(dioxane)_x]_\infty$, where (a) R = I, x = 5 form 1D linear chains, (b) R = Br, x = 4 from (4,4) 2D square nets, and (c) R = F and x = 3 form 3D cubic lattices, **Figure 8** [51]. This remarkable set of complexes again demonstrates that the desired hexameric $Na_6O_6$ SBU is robust and remains intact within each network. Also, the type of network can again be rationalized in terms of the size of the substituent at the *para* position, with only the fluorine group being small enough to fit into the available volume within the cubic cavities. Moreover, we have again confirmed that network formation using the $Na_6O_6$ SBUs is limited to these three structural types, *i.e.* structural characterization of the dioxane solvated *para*-aryloxides R-$C_6H_4ONa$: R = Cl, I or Ph, give 1D linear chains, R = Br, H, Me, $^i$Pr, $^t$Bu or $^t$Pn give 2D (4,4) nets and R = F or Et give 3D cubic networks. These results show that even the heavier *s*-block elements may be used as robust SBUs.

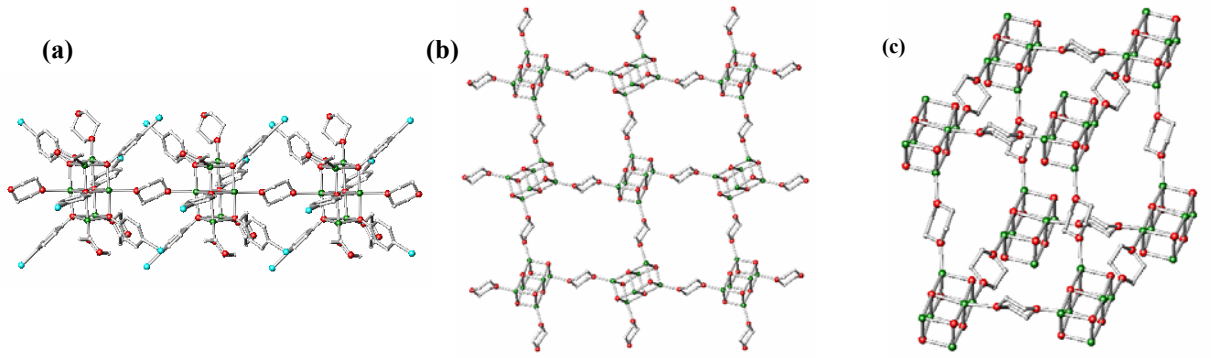



**Figure 8.** Sections of the polymeric structures of the: (a) 1D linear chain, **(b)** 2D square net (framework atoms only), and **(c)** 3D cubic lattice (framework atoms showing a single cubic unit), formed from triple-stack $Na_6O_6$ SBUs bridged by dioxane.

## 5. CONCLUSIONS

To summarize, we have demonstrated that both lithiated and sodiated pre-assembled molecular aggregates may be used to construct network assemblies. The topologies of the resulting frameworks are directed by the geometry of the metals within the molecular SBUs and is also dependant on the ability of organic components to fill space effectively. These initial studies suggest that there is a rich chemistry surrounding the use of early main group metals in the supramolecular synthesis of extended frameworks.

## 6. ACKNOWLEDGEMENTS

We gratefully acknowledge the Petroleum Research Fund (41716-AC3) and the University of Notre Dame for support. We also thank the National Science Foundation for instrumentation support (CHE-0443233).

## 7. REFERENCES

[1] L. Brammer, *Chem. Soc. Rev.* **33** (2004) 476.
[2] D. Braga, *Chem. Commun.* (2003) 2751.
[3] S. L. James, *Chem. Soc. Rev.* **32** (2003) 276.
[4] C. Janiak, *Dalton Trans.* (2003) 2781.
[5] J. K. M. Sanders, *Chem. Eur. J.* **4** (1998) 1378.
[6] G. B. Gardner, D. Venkataraman, J. S. Moore and S. Lee, *Nature* **374** (1995) 792.
[7] S. J. Lee and W. Lin, *J. Am. Chem. Soc.* **124** (2002) 4554.
[8] H. O. Stumpf, L. Ouahab, Y. Pei, D. Grandjean and O. Kahn, *Science* **261 (**1993) 447.
[9] J. W. Steed and J. L. Atwood, *Supramolecular Chemistry*, (Wiley, New York, 2000).
[10] G. R. Desiraju, *Crystal Engineering. The Design of Organic Solids*, (Elsevier, Amsterdam, 1989).
[11] I. Haiduc and F. T. Edelmann, *Supramolecular Organometallic Chemistry*, (Wiley-VCH, New York, 1999).
[12] V. Chandrasekhar, S. Nagendran, S. Bansal, A. W. Cordes and A. Vij, *Organometallics* **21** (2002) 3297.
[13] O. M. Yaghi, M. O'Keeffe, N. W. Ockwig, H. K. Chae, M. Eddaoudi and J. Kim, *Nature* **423** (2003) 705.
[14] T. Stey and D. Stalke, D. in: *The Chemistry of Organolithium Compounds*, eds. Z. Rappoport and S. Patai (Wiley, New York, 2004), p 47.
[15] A. M. Sapse and P. v. R. Schleyer, *Lithium Chemistry, A Theoretical and Experimental Overview*, (Wiley, New York, 1995).
[16] E. Weiss, *Angew Chem. Int. Ed. Engl.* **32** (1993) 1501.
[17] An excellent special addition of the Journal of Solid-State Chemistry is available: O. M. Yaghi, M. O'Keefe, eds. *J. Solid State Chem.* **152** (2000) 1-321.
[18] K. W. Henderson, A. R. Kennedy, A. E. McKeown and D. S. Strachan, *J. Chem. Soc., Dalton Trans.* (2000) 4348.
[19] S. E. Denmark, K. A. Swiss and S. R. Wilson, *Angew. Chem. Int. Ed. Eng.* **35** (1996) 2515.
[20] A. F. Wells, *Three-Dimensional Nets and Polyhedra*, (Wiley-Interscience, New York, 1977).
[21] J. Kim, B. Chen, T. M. Reineke, H. Li, M. Eddaoudi, D. B. Moler, M. O'Keeffe and O. M. Yaghi, *J. Am. Chem. Soc*. **123 (**2001) 8239.
[22] G. Boche, *Angew. Chem. Int. Ed. Engl.* **28** (1989) 277.
[23] K. W. Henderson, A. R. Kennedy, L. Macdonald and D. J. MacDougall, *Inorg. Chem.* **42** (2003) 2839.
[24] K. W. Henderson, A. R. Kennedy and D. J. MacDougall, *Inorg. Chem.* **42** (2003) 2736.
[25] K. W. Henderson, A. R. Kennedy and D. J. MacDougall, *Phosphorous, Sulfur and Silicon* **179** (2004) 795.
[26] N. W. Ockwig, O. Delgado-Friedrichs, M. O'Keeffe and O. M. Yaghi, *Acc. Chem. Res.* **38** (2005) 176.
[27] D. Braga, F. Grepioni and G. R. Desiraju, *Chem. Rev.* **98** (1998) 1375.
[28] B. Walfort, L. Lameyer, W. Weiss, R. Herbst-Irmer, R. Bertermann, J. Rocha and D. Stalke, *Chem. Eur. J.* **7** (2001) 1417.
[29] R. E. Dinnebier, U. Behrens and O. Falk, *J. Am. Chem. Soc.* **120** (1998) 1430.
[30] For a specific example of 1,4-dioxane bridging in a lithiated coordination polymer see: O. Kühl, J. Sieler, G. Baum and E. M. Hey-Hawkins, *Z. Anorg. Allg. Chem.* **626** (2000) 605.
[31] T. J. Boyle, D. M. Pedrotty, T. M. Alam, S. C. Vick and M. A. Rodriguez, *Inorg. Chem*. **39** (2000) 5133.
[32] K. Thiele, H. Gorls and W. Seidel, *Z. Anorg. Allg. Chem*. **624** (1998) 1391.
[33] L. M. Jackman, D. Cizmeciyan, P. G. Williard and M. A. Nichols, *J. Am. Chem. Soc.* **115** (1993) 6262.
[34] P. K. Chattaraj, H. Lee and R. G. Parr, *J. Am. Chem. Soc.* **113** (1991) 1855.
[35] F. T. Deelmann, F. Pauer, M. Wedler and D. Stalke, *Inorg. Chem.* **31** (1992) 4143.
[36] M. Motevalli, D. Shah and A. C. Sullivan, *J. Chem. Soc, Dalton Trans*. (1993) 2849.
[37] R. Robson, in: *Comprehensive Supramolecular Chemistry*, eds. J. L. Atwood, J. E. D. Davies, D. D. MacNicol, F. Vögtle and K. S. Suslick, (Elsevier, Oxford, 1996).
[38] F. H. Allen, *Acta Crystallogr.* **B58 (**2002) 380.
[39] For example see: B. W. Skelton, C. R. Whitaker and A. H. White, *Aust. J. Chem.* **43** (1990) 755.

[40] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, C. G. Zakrzewski, J. A. Montgomery Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman [illegible]7[illegible]cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, [illegible] Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, J. L. Andres, M. Head-Gordon, E. S. Replogle and J. A. Pople, *Gaussian 98*, revision A.6; Gaussian, Inc.: Pittsburgh, PA, 1998.
[41] W. J. Hehre, R. Ditchfield and J. A. Pople, *J. Chem. Phys.* **56** (1972) 2257.
[42] P. C. Hariharan and J. A. Pople, *Theo. Chem. Acta* **28** (1973) 213.
[43] A. D. Becke, *J. Chem. Phys.* **98** (1993) 5648.
[44] A. D. McLean and G. S. Chandler, *J. Chem. Phys.* **72** (1980) 5639.
[45] L. M. Jackman and B. D. Smith, *J. Am. Chem. Soc.* **110** (1988) 3829.
[46] D. J. MacDougall, J. J. Morris, B. C. Noll and K. W. Henderson, *Chem. Commun.* (2004) 456.
[47] S. Mecozzi and J. Rebek Jr., *Chem. Eur. J.* **4** (1998) 1016.
[48] R. E. Dinnebier, M. Pink, J. Sieler and P. W. Stephens, *Inorg. Chem.* **36** (1997) 3398.
[49] T. J. Boyle, N. L. Andrews, M. A. Rodriguez, C. Campana and T. Yiu, *Inorg. Chem.* **42** (2003) 5357.
[50] M. Kunert, E. Dinjus, M. Nauck and J. Sieler, *Chem. Ber.* **130** (1997) 1461.
[51] D. J. MacDougall, B. C. Noll and K. W. Henderson, *Inorg. Chem.* **44** (2005) 1181.

# MODELING STUDIES OF ANTI-AMYLOID ANTIBODIES BOUND TO Aβ

Anna S. Gardberg, Nathan Tanner, Chris Dealwis*
*University of Tennessee at Knoxville, Department of Biochemistry and Molecular Biology, M407 Walters Life Sciences*
*Knoxville, Tennessee 37996-0830*

Israel Huff, Kimberly Salone, Ronald Wetzel
*University of Tennessee at Knoxville, Graduate School of Medicine, Knoxville, Tennessee, USA*
*Corresponding author: cdealwis@utk.edu
Keywords: A-beta, beta-amyloid, monoclonal antibodies, Alzheimer's, neurodegenerative disorders Autodock, protein-protein interaction
Grant sponsors: AG, NT, IH, RW, and CD are supported by NIH grant 5R01NS046356-03. KS was supported by a grant from the American Health Assistance Foundation.

## 1. ABSTRACT

WO1 and WO2 are conformation-specific monoclonal IgMs that bind the fibril state of the amyloid Aβ peptide (1-40), as well as amyloid fibrils of other disease-related proteins. Significantly, these antibodies (Abs) do not bind the soluble, monomeric state of Aβ (1-40) or the precursor form of other amyloids. The Abs have been sequenced and compared with sequences in the Kabat database; though some unusual charged residues were noted in WO1 and WO2, none are at the CDR. Three-dimensional models of the Fv fragments of WO1 and WO2 were generated with Web Antibody Modeling. A trigonal prism and a two-strand structural model of the Aβ amyloid core were compared by docking each with the Fv models of WO1 and WO2. The results predict binding of WO1 and WO2 to the Aβ trigonal prism at the protofilament face consisting of residues 23-27 and of WO2 to the Aβ two-strand at the end of the elongating fibril. No positive prediction for WO1 and the Aβ two-strand was made. Binding experiments with WO1 were determined to be sensitive to salt and pH conditions, suggesting the importance of electrostatic interactions for binding; this result is consistent with the proposed docking of WO1 to the Aβ trigonal prism.

## 2. INTRODUCTION

The aggregation of normally soluble proteins into insoluble, unbranched fibrils is the underlying pathology of a family of diseases known as the amyloidoses.[1] The hallmark event in amyloidogenesis is a change in the secondary and/or tertiary structure of a normal, soluble protein, rendering it prone to self-assembly into highly ordered para-crystalline arrays: fibrils. More than 20 proteins have been clinically identified as precursors of amyloid fibrils *in vivo*. These include the amyloid precursor protein (APP), Islet amyloid polypeptide (IAPP), α-synuclein, transthyretin (TTR), immunoglobulin light chain (LC), polyglutamine-repeats, and prion proteins, that are associated with diseases such as Alzheimer's,[2-4] type II diabetes,[5,6] Parkinson's disease,[7-9] familial polyneuropathy,[10] light chain associated (AL) amyloidosis,[11-13] Huntington's disease,[4,14] and the spongiform encephalopathies.[15] By understanding the three-dimensional structure of such fibrils, we might design therapeutic agents to target them. It is notoriously difficult to extract structural information directly from amyloid fibrils, which are insoluble and non-crystalline, so a great deal of study has gone into fiber diffraction studies,[16-18] mutation studies,[19] microscopy,[17,20,21] proteolysis,[22,23] ESR,[24,25] NMR,[24,26] SANS,[27] and deuterium exchange with mass spectroscopy[28-31] in an ongoing attempt to elucidate the molecular structure of the Aβ fibrils associated with Alzheimer's disease.

There is no consensus model for Aβ structure, but most models incorporate cross-beta secondary structure. The cross-beta structure (in which the fibril axis is perpendicular to the chain direction) of Aβ models is derived from the 4.75 Å meridional reflection in fiber diffraction studies.[18] A stacked, parallel beta-sheet arrangement is suggested by Burkoth, Benzinger, and others,[24,27,32] while a pair of concentric cylinders was proposed by Perutz, *et al*.[33] An anti-parallel or stacked-hairpin beta-sheet arrangement is suggested by Callaway and others.[34-37]

One recent model of the Aβ protofilament, proposed by Guo, *et al,*[38] features a trigonal prism of stacked, parallel Aβ(15-36) polypeptides; the model is shown in **Figure 1a**. This trigonal prism protofilament model is consistent with threading analysis[38] and proline-scanning mutagenesis,[19] and the proposed fibril model (of 6 protofilaments) appears to be consistent with electron microscopy. Recent data from hydrogen exchange NMR[39] and cysteine scanning mutagenesis[40] have called portions of this construction into question.

Another proposed configuration of the Aβ protofilament consists of two stacks of two parallel strands twisting around one another along the fibril axis; this model, shown in **Figure 1b**, is consistent with data that was available in 2002, especially that from solid-state NMR studies.[26] It is also in line with the recent crystal structure of the amyloid form of a seven-residue fragment from yeast protein Sup35[41], which supports a double-beta-sheet structure. Both the two-strand and trigonal prism Aβ protofilament constructs are consistent with specific bodies of experimental data, but are not completely determined by experimental data. Both models continue to evolve with the acquisition of new data.

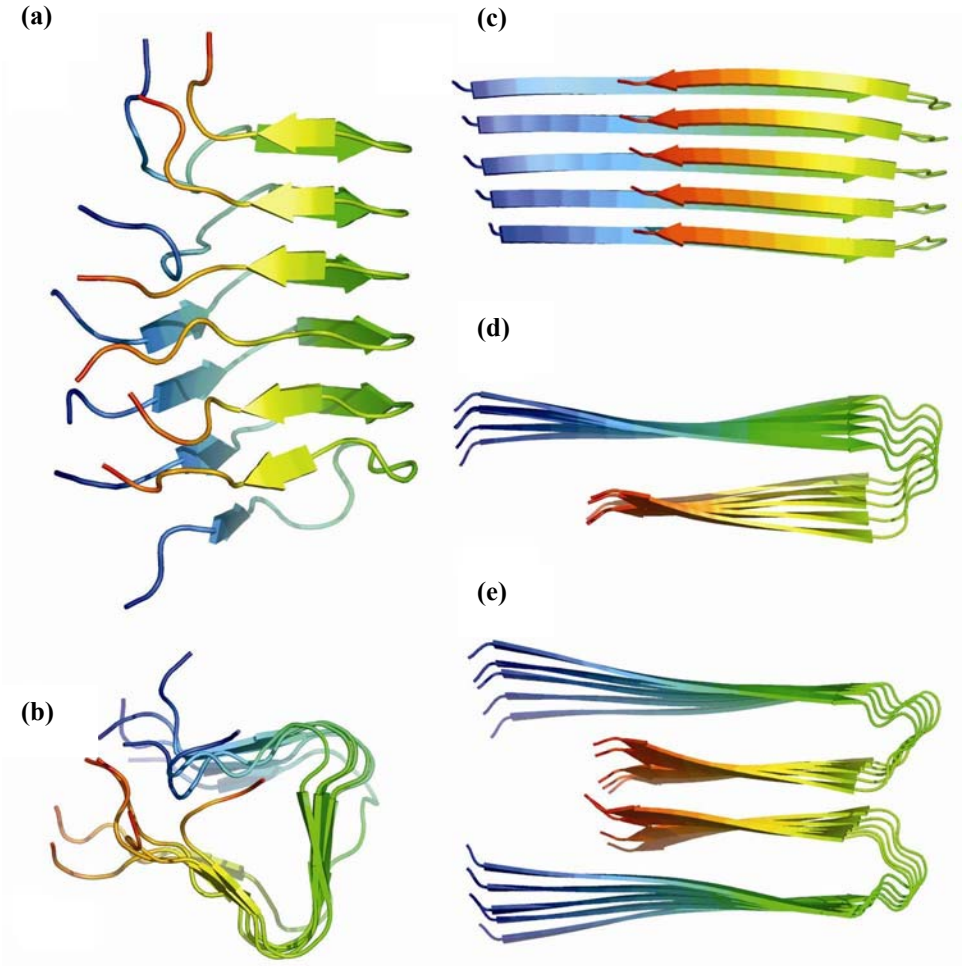



**Figure 1. (a)** Three-strand protofilament model in side view. **(b)** Three-strand, top view. **(c)** One half of proposed two-strand Aβ, side view. **(d)** Top view. **(e)** Two-fold symmetry about the *x*-axis has been used to generate one proposed configuration for the full protofilament from the two-strand model.

The monoclonal antibodies WO1 and WO2 (κ-light chain, murine, IgM) recognize a common conformational epitope shared by several different types of amyloid fibrils, with little dependence on amino acid sequence; the two bind the same antigen with similar affinities for the Aβ fibril.[42] This result opens the door to targeting whole classes of amyloid fibrils for detection and therapy.[43] Work to crystallize these antibodies (or their fragments) is ongoing. While IgM's are notoriously difficult to model,[44] the success of antibody modeling algorithms at predicting crystal structures of antibody variable regions[45] encouraged the use of computational models of WO1 and WO2 Fvs for docking simulations. Since WO1 and WO2 bind to Aβ amyloid, we hypothesized that a model of

the Aβ protofilament would, if correct, dock *in silico* to a model of the variable regions of WO1 and WO2. The configuration of the predicted complexes of the Fvs of antibodies WO1 and WO2 to two different models of Aβ protofilament is given. Experimental salt-dependence data is consistent with the prediction of WO1-Aβ binding for the trigonal prism model.

## 3. METHODS

### 3.1 Sequence Analysis

The nucleotide sequences of WO1 and WO2 were obtained by cloning and confirmed by multiple cloning using high fidelity polymerases, and further confirmed by extended N-terminal amino acid sequencing from analysis of the protein. Initial alignments were performed using Molecular Operating Environment (MOE™[46]), with Ab sequences obtained from the Protein Data Bank[47,48]. An extensive alignment was done using the Kabat sequence database testing program[49]. BLASTP analysis[50] was run on the WO1 and WO2 variable light chains using the Non-Redundant Protein Database at the San Diego Supercomputer Center. Canonical classes are based on Chothia nomenclature[51] and numbering scheme is that of Kabat[52]. The WO1 and WO2 Fv residue sequences were initially compared to one another. Then they were compared with sequences from the Kabat database to search for anomalies, unique features, or homologies to other antibodies. A sequence alignment of the CDRs is shown in Figure 2.

### 3.2 Model Building

Three-dimensional structures of WO1 and WO2 variable regions (Fv) were generated using the Web Antibody Modeling algorithm, WAM.[45] WAM is an improvement on the AbM program [53-56] with greater capability in modeling the highly variable H3 loop through a combination of knowledge-based and *ab initio* methods[45] . The WAM algorithm has a record of producing models that are 1.0-2.8 Å RMSD from observed structures for the heavy chain CDR3, and better than that for the canonical loops (typically 1.0-2.5 Å RMSD).[45]

### 3.3 Docking

#### 3.3.1 Preparation

A PDB file of the trigonal prism Aβ 1-40 hexameric protofilament model after molecular dynamics simulations ("Aβ trigonal prism", which models residues Gln15 through Val36)[38] was kindly provided by Juntao Guo and Ying Xu. A PDB file of the two-strand pentameric Aβ 1-40 protofilament model after energy minimization ("Aβ two-strand", which models residues 9 through 40)[26] was kindly provided by Robert Tycko.

The program Autodock3[57] predicts the interaction of ligands with macromolecular targets. As Autodock3 can handle no more than 2048 atoms, we reduced the two-strand protofilament to a tetramer and did not generate its symmetry pair. Control calculations performed with a protein-peptide complex, a protein-ligand complex, and a lysozyme-anti-lysozyme complex of known structure (PDB codes 1SMR, 1DQJ and 1DYI) showed that the charges assigned from forcefield calculations performed by AutoDockTools yielded less accurate results than those assigned by AMBER forcefield calculations[58,59] performed by InsightII.[60] We assigned charges using the following protocol. The atom names were standardized to Refmac5 conventions. Using Refmac5[61,62], hydrogen atoms were added in “riding” positions computed from the carbon, nitrogen, oxygen, and sulfur atomic positions. After capping the termini of each chain with (neutral) carboxylic and amine end groups, the partial charges for each atom (including polar hydrogen atoms) in each model were calculated with the InsightII program using the AMBER force field. Calculating the summed formal and partial charges for each model verified that the model is electrically neutral. A Sybyl-style file was then exported from InsightII.

#### 3.3.2 Computation

Autodock-style PDBQ files were prepared from the Sybyl-style files (described in the previous section) with *mol2topdbq* and *mol2topdbqs*, awk-based utilities packaged with Autodock3.[57] For purposes of solvation calculation, we designated the Fv models the “proteins” and the Aβ protofilament models the “ligands”. Despite the thermodynamic importance of side-chain motion,[63] neither the Fvs nor the Aβ models were permitted any torsional freedom; the Autodock package does not permit the side chains of the protein to move, and the Autotors utility for designating rotatable bonds in the ligand allows a maximum of 32, not nearly sufficient for a the 2395-2040-atom Aβ models. Grid parameter files were generated *via mkgpf3* (a script which comes packaged with

Autodock3) and adjusted to compensate for the large volume required for the interaction of the Fab's complementarity determining region (CDR) with the protofilament models. The grid maps were calculated using AutoGrid. The gridded area dimensions were 66 Å x 66 Å x 66 Å, with grid spacing increased to 0.55 Å. The final gridded regions encompassed the CDR of the Fv (as well as all of the heavy chain and most of the light chain of the model for WO1, and all save the C-termini for WO2) and the starting position of the Aβ model, with room for rotation and translation. Grid generation was performed with Autogrid3.
Docking was performed using the Lamarckian genetic algorithm (LGA), and the pseudo-Solis and Wets methods were applied for the local search. Each docking experiment was performed 50 times. Default parameters were applied for the docking experiments, with the exception of the initial rotation step size, which was set to 180. The results of the docking experiments were evaluated by calculating the positional root-mean-square (rms) deviation of the corresponding atoms of each conformation. Docking parameter files were prepared with the *mkdpf3* setup utility and adjusted by hand.

### 3.3.3 Contact Analysis

Contacts between the Fv models and the protofilament models in the predicted docking complexes were analyzed with the program CONTACT.[62] Docking clusters and configurations were examined Pymol,[64] with residue charge surfaces visualized. Electrostatic surface plots for publication were generated by the DelPhi module of InsightII.[60] Angles between solutions were computed by calculating the average direction vector along the protofilament axis for each solution and taking the inverse cosine of the dot-product of direction vectors.

### 3.4 Electrostatic Analysis

Electrostatic surfaces were calculated using DelPhi, a part of the InsightII graphical software package. The following parameters were employed: solvent dielectric 80, radius 1.4, ion radious 2.0, grid solvent extent 15, 65 points. Models were displayed and manipulated with InsightII and all computational work done on a Silicon Graphics computer system.

### 3.5 Salt Effects on Binding WO1

Binding of the WO1 antibody to Aβ fibrils was tested under 13 different conditions of cation, anion, salt concentration, and pH. 1xPBS buffer was used as a control. Experiments 1-5 comprised the salt-effects experiments; they consisted of: 5 mM HEPES at pH 7.5 with: 1) no salt, 2) 150 mM NaCl, 3) 600 mM NaCl, 4) 150 mM KCl, and 5) 75 mM $Na_2HPO_4$. In experiments 6-13, the effects of pH were examined; they consisted of:150 mM NaCl with 6) citric acid pH 3.0, 7) citric acid pH 5.8, 8) PIPES at pH 5.8, 9) PIPES at pH 7.4, 10) Bicine at pH 7.4, 11) Bicine at pH 8.5, 12) CHES at pH 8.5, and 13) CHES at pH 10.0.

No ionic interference was expected from HEPES, which has low ionic strength. $Na_2HPO_4$ has roughly double the ionic strength of the other two salts so it was used at half the concentration for a balanced comparison. The salt-effect experiments were buffered to pH 7.5, near physiological pH conditions. The pH effect experiments contained 150 mM NaCl to provide approximate physiological ionic strength conditions. The Aβ fibrils were prepared as described in *Kheterpal, et al, 2000*.[31]

## 4. RESULTS

### 4.1 Sequence Analysis

### 4.1.1 Comparison of WO1 to WO

The light chain variable regions share 66% identity (80% similarity) and the heavy chain variable regions share 82% identity (85% similarity). Overall charge composition of the Fv fragments is very similar, with isoelectric points (pIs) of the variable regions calculated to be: WO1 $V_H$ 7.3, $V_L$ 8.7, WO2 $V_H$ 7.2, and $V_L$ 8.7.

While both sequences are rich in hydrogen bonding residues, WO1 $V_L$ CDR has a greater proportion of charged Asp, Glu, Arg, and Lys residues (19% vs. 11%), while the WO2 $V_L$ CDR is richer in uncharged, polar hydroxyl residues (42% for WO1 $V_L$, 64% for WO2 $V_L$). The sequences of the CDR's are compared in **Figure 2**. The WO1 and WO2 heavy chain variable regions are more similar in overall sequence, and the number of charged residues in the $V_H$ is identical. Still, the two charged residues of WO1's $V_H$ are at its CDR, and those of WO2's $V_H$ are not at its CDR. WO1's light chain CDR contains charged residues at positions occupied by hydroxyl residues in WO2 such as Arg24, Lys52, and Asp56 in WO1 which are replaced by Thr, Ser, and Ser in WO2. Similarly, the unique charged

residue Arg93 in WO2 replaces a serine residue in WO1. WO1's greater proportion of charged residues and WO2's greater proportion of uncharged polar residues suggest that WO1 and WO2 bind amyloid differently from one another.

a)

| VL CDR1 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 31a | 32 | 33 | 34 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| WO1 | R | A | S | G | N | I | H | N | - | ***Y*** | L | A |
| WO2 | T | A | S | S | S | V | S | S | S | ***Y*** | L | H |
| α-Lyso | R | A | S | Q | N | I | S,H | N | N | - | L | H,A |
| Germ. | R,K | S,A | S | Q | S,N | I | nh* | nh | S | nh | L | nh |
| Kappa | R,K | A,T | S | Q | nh | V,I | nh | S,T | - | nh | L | nh |

| CDR2,3 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| WO1 | N | A | K | T | L | ***A*** | D | Q | H | F | W | S | T | P | ***Y*** | T |
| WO2 | S | T | S | N | L | ***A*** | S | L | Q | Y | H | R | S | P | ***Y*** | T |
| α-Lyso | Y | T | T,S | Q,T | S,L | nh | D,S | Q | Q | F,Y | nh | S | - | P | R,L | T |
| Germ | nh | A,V | S,A | nh | R,L | D,E | S | nh | Q | nh | S | nh | nh | P | R | T |
| Kappa | nh | A | S | nh | R,L | nh | S | Q | Q | S,Y | nh | S,D | nh | P | R,L | T |

b)

| VH CDR1,2 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| WO1 | G | Y | T | F | T | E | Y | ***T*** | M | H | G | I | ***N*** | P | N | ***N*** | G | T | **S** |
| WO2 | G | Y | S | F | T | G | Y | ***T*** | M | N | L | I | ***N*** | P | Y | ***N*** | G | T | **S** |
| α-Lyso | G | Y | T,S | F | T | T | Y | W | I,V | E,S | E | I | L | P | G | S | nh | T | Y,D |
| Germ | G | F | T | F | T | D | Y | Y | M | S,H | F | I | R | nh | nh | nh | nh | T | E,K |
| Kappa | G | F,Y | T | F | T | D | Y | Y | I,M | H,K | nh | I | nh | P | nh | nh | G,S | T | nh |

| CDR3 | 96 | 97 | 98 | 99 | 100 | 101 | 102 | 103 | 104 | 105 | 106 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| WO1 | C | A | R | D | D | G | - | ***Y*** | Y | G | F |
| WO2 | C | A | R | R | L | G | D | ***Y*** | Y | A | M |
| α-Lys | C | A | R,S | nh | D,E | G | - | nh | Y | G | nh |
| Germl | C | A | R,S | D | nh | nh | nh | nh | Y | A,Y | M,G |
| Kappa | C | A | nh | G,D | nh | G | nh | nh | nh | nh | F |

**Figure 2.** CDR composition of a) $V_L$s and b) $V_H$s compared with those of anti-lysozyme, germline, and kappa-light-chain antibodies. **nh* signifies no homology. Bold, italicized text highlights conserved residues in WO1 and WO2 White-on-black text indicates identical or homologous residues for WO1, WO2, and one or more compared sequences

### 4.1.2 Light Chain

A preliminary sequence alignment was performed using a database of various antibody sequences, including subsets of anti-lysozyme Ab chains and germline Ab chains. The results showed a high degree of homology between heavy and light chain sequences, with several residues noted as distinctive in **Table 1**. The CDRs of the light and heavy chains of WO1 and WO2 are rich in hydrogen-bonding residues relative to anti-lysozyme and germline antibodies, suggesting that hydrogen bonding plays an important role in amyloid-recognition.

A more extensive alignment was done using the Kabat sequence database testing program. The WO1 $V_H$ and $V_L$ sequences were compared to the database of 2707 light chains and 3471 heavy chains. The unique residues Lys42 and Lys74 had no consensus within the database; at these positions, which are not at the CDR, Lys occurred in less than 1% of the sequences. Lys52 at light chain CDR2, however, was a hydroxyl residue in the majority of the 2701 chains. None of the WO2-unique charged or hydroxyl residues showed any incongruity.

| Light Chain Residue WO1 WO2 | Equivalents | Residue exhibited by chains with non-identity |
|---|---|---|
| Lys42 Ser43 | No | None charged, only Gly, Gln, Tyr |
| Lys52 Ser53 | Lys only in WO1 | No charged, 95% Tyr or Ser |
| Asp56 Ser57 | Asp in WO1 and (AL) | 99% Pro, (G) have Thr/Ser |
| Lys74 Thr75 | Lys only inWO1 | 100% Glu/Asp, (G) has Thr |

**Table 1:** Light chain-unique residues and comparison for WO1 residues that are less than 10% homologous to database sequences in the initial alignment tests. The alignment was done with specific interest in the anti-lysozyme Abs as well as germline Abs, and any residues here are noted if appearing in either chain. The residue is listed with its location, as well as the residues that were most commonly shared throughout the other Ab sequences. The equivalent residue from WO2 is also included for comparison. Codes: antilysozyme (AL), germline (G).

### 4.1.3 Heavy Chain

Through the preliminary alignment including the anti-lysozyme, germline, and kappa-light chain sequences, several residues were noted as distinctive. They are listed in **Table 2**. As with the light chain results, the most striking disparities between the WO1/WO2 heavy chains and comparison germline and anti-lysozyme sequences occur with charged residues. Again, the Kabat alignment was performed. The equivalent positions for residues Lys63, Lys65, Lys67, Lys74, Asp73 (in WO1 and WO2) and Arg84 (in WO1 only) have no consensus, but these charged residues are not unique and do not occupy the CDR. The only significant discrepancy was the Asn44 of the WO2 heavy chain. The Asn residue found at this site is found in only 13 examples, 0.326% of the database. Typically, a Ser residue is at this position. This position is also not at the CDR.

| Heavy Chain Residue WO1 WO2 | Equivalents | Residue exhibited by chains with non-identity |
|---|---|---|
| Lys63 Lys63 | WO1/WO2, (G) | No other charged residues |
| Lys65 Lys65 | WO1/WO2, (G), (AL) | 100% Val, Leu, or Phe |
| Lys67 Lys67 | WO1/WO2, (G), (AL) | 100% Gly, Ser, Asp |
| Asp73 Asp73 | WO1/WO2, (G), (AL) | Mostly Asp/Glu |
| Lys74 Lys74 | WO1/WO2, (G) | 99% Asp/Glu |
| Arg84 Leu84 | Arg only in WO1 | Non-consensus, but no other charged |

**Table 2:** Heavy chain unique residues and comparison. The residues displaying less than 10% homology in the initial alignments to database sequences are included here. Any germline or anti-lysozyme Ab commonalities are noted, as well as the residues which occur in the chains which are not homologous to the WO1/WO2 heavy chains. Codes: antilysozyme (AL), germline (G).

### 4.1.4 Fv Models

Of the WO1 light chain's unique lysine residues (42, 52, and 74), only Lys52 is at the CDR and may play a role in docking. However WO1 light-chain Asp56 (unique to WO1 and anti-lysozyme Abs) may play a role, since it makes a hydrogen bond to an Asn residue in the trigonal prism model.

### 4.2 Electrostatic Surface Potentials for WO1 and WO2

WO1 and WO2 Fvs display regions of significant charge density (as calculated by the DelPhi module of InsightII), observed in the electrostatic potential map shown in Figure 3. Though the pIs of the WO1 and WO2 Fvs are nearly neutral, they clearly have highly charged regions. The presence and distribution of these charged segments indicates a possible role of charge interactions in antigen binding. The line of positively charged residues on each of the light-chains suggests a template for hydrogen bonding and salt bridges. The line of negatively charged residues at the heavy-light interface of WO1's CDR (as seen in **Figure 3**) reflects the higher proportion of negatively charged residues in its composition.

### 4.3 Electrostatic Surface Potentials for the Protofilament Models

The most striking features of both models are the bands of charged residues aligned along the stacking axis; these representations are shown in **Figure 4**. The charge bands are a consequence of the in-register parallel-stacking of the models. Labeling the trigonal prism protofilament model's face containing residues 17-22 "A", 23-27 "B", and 28-36 "C" (Aβ 1-40 numbering basis), we see that faces A and C display exposed hydrophobic residues. In the two-strand model, the N-terminal face consists of residues 9-23; this corresponds to the "A" face of the trigonal prism

model. There is then a 180º turn consisting of residues 24-27; this turn (instead of a β sheet) is in sharp contrast to the "B" face of the other model. Finally, residues 30-40 form the C-terminal face of the two-strand model.

**(a)** **(b)**

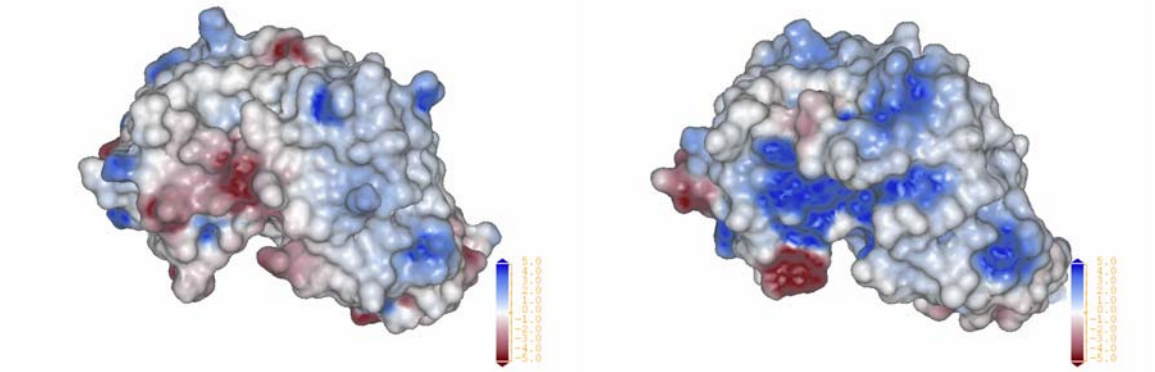


**Figure 3**. Electrostatic surfaces for WO1 **(a)** and WO2 **(b)** Fv models. The viewer faces the CDRs. In each image, the heavy chain fragment is on the left.

**(a)**

**(b)**

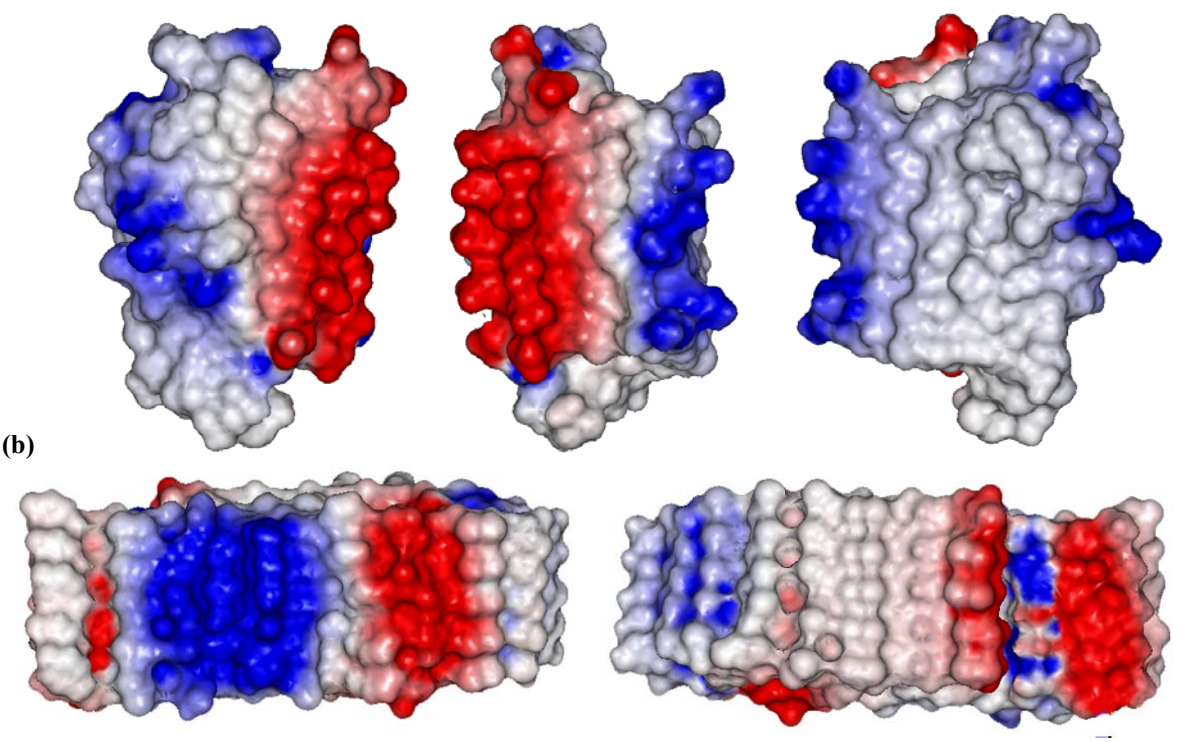


**Figure 4**. **(a)** Electrostatic surfaces for the A, B, and C faces of the trigonal prism protofilament model. The line of Glu22 and Asp23 residues (shown in red) at the corner of the A and B faces is particularly striking. **(b)** Electrostatic surfaces for the two faces of the two-strand model. The N-terminal face is shown in the left-hand image; strand direction is left to right. The C-terminal face is shown in the right-hand image; strand direction is left to right.

### 4.4 Docking WO1-Trigonal Prism-Model

AutoDock3 predicted 50 binding configurations. Both the Fv and the trigonal prism model were treated as rigid bodies. Autodock3 found 41 distinct conformational clusters (using an RMSD-tolerance of 1.0 Å), of which 8 contained more than one solution. The variation in energy from the best to the worst solution was 7.0 kcal/mol.

Although there was one cluster of two solutions in the top 7 (with the B face laterally at the heavy-light interface), more intriguing is a collection of related clusters beginning at the 8th-ranked solution. In total, 21 solutions have the B face of the trigonal prism protofilament model longitudinally oriented at the light-chain CDR, with some overlap onto heavy chain CDR. Some are offset by one, two, or three monomers (i.e., the stack of six Aβ monomers is shifted along the stacking axis), others by angles less than 27˚. This is the largest supercluster; its lowest-energy solution is presented in **Figure 5a**. The particular solution used for the following analysis was the lowest-energy member of this supercluster, which is 3.5 kcal/mol (~1 hydrogen bond) less favorable than the lowest-energy solution of all docking runs.

### 4.4.1 Salt Bridges and Ion Pairs

This docking model indicates that the line of Glu22 and Asp23 residues in the trigonal prism model binds to the corresponding line of polar and positively charged residues of WO1's light-chain CDR through salt bridges and type 2 hydrogen bonds between acidic residues and carbonyl groups. The line of Lys28 residues likewise interact with the trail of polar and negatively charged residues of the light *and* heavy chain CDRs. Breaking the ion pairs down by Szilyagi and Zavodsky classification,[65] there are four bonds at < 4.0 Å, six bonds between 4.0 and 6.0 Å, and ten pairs between 6.0 and 8.0 Å. *Hydrogen bonding*. The O and N atoms of the side chains of the line of Asn27 residues on the three-strand model make hydrogen bonds with Tyr and Asp residues of WO1. The carbonyl O atoms of the line of (inward-pointing) Val24 and Gly25 residues bond to Tyr residues of WO1. The backbone N of inward-pointing Ser26 makes some possible hydrogen-bonds. There are 17 likely hydrogen bonds at less than 3.5 Å, two of which were also counted as close ion pairs. The hydrogen bonding interface is shown in **Figure 5d**.

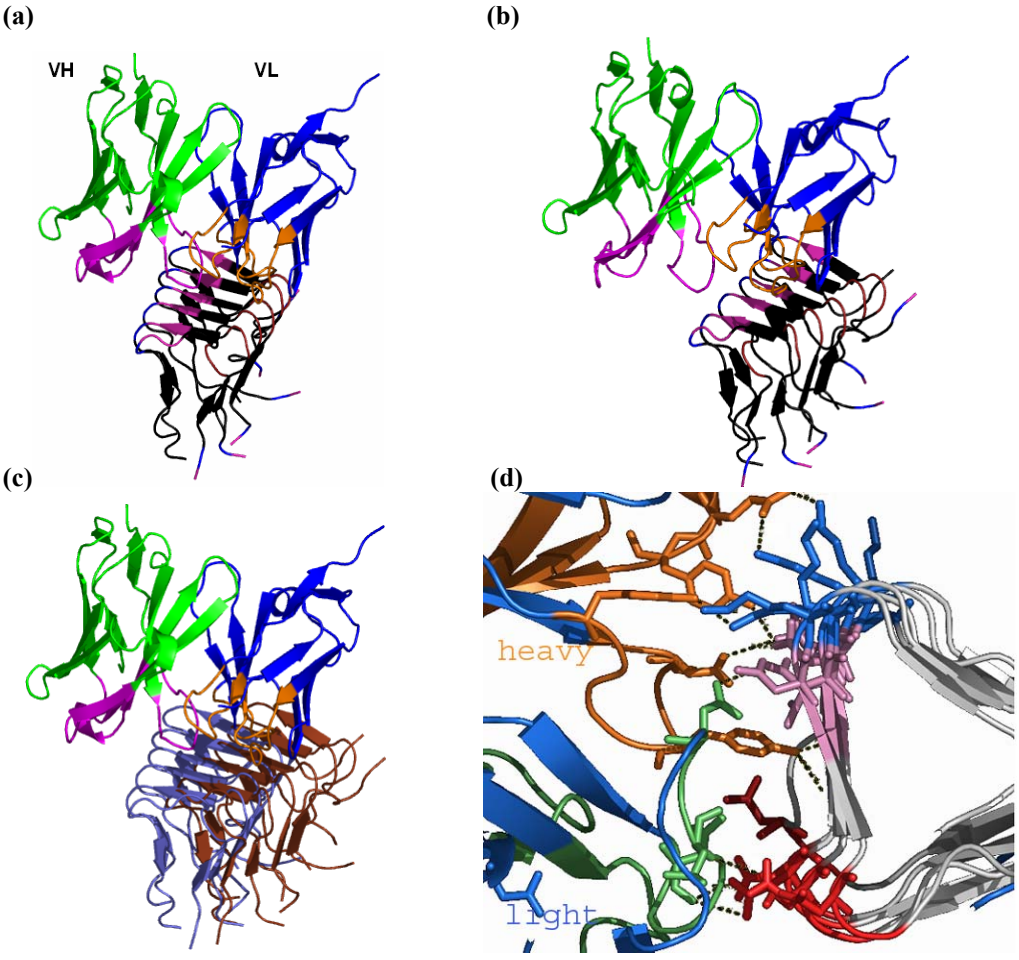



**Figure 5**. **(a)** WO1 and **(b)** WO2 Fv docked with trigonal prism protofilament model in the predicted configuration (top solution in supercluster). The heavy chain is on the left, the light chain is on the right, and the CDRs are highlighted. The trigonal prism model is shown colored according to residue: negative, red; positive, blue; polar, pink; hydrophobic, black. **(c)** The two solutions are showed overlain. **(d)** A close-up view of the docking interface for the WO1-trigonal-prism model. The color code is as for **(a)** and **(b)**, but hydrophobic residues are shown in grey and hydrogen bonds are shown in black.

**4.4.2 Hydrophobic Interactions**
There are no aromatic rings on the B-face of the trigonal prism model, and the hydrophobic Val24 side chain points toward the interior of the trigonal prism model. Thus hydrophobic interactions between WO1 and the trigonal prism model are not observed in this docking model.

**4.5 Docking WO2-Trigonal Prism-Model**
AutoDock3 predicted 50 binding solutions. The Fv and the trigonal prism model were both treated as rigid bodies. For WO2 and the trigonal prism protofilament model, Autodock3 found 42 distinct conformational clusters (using an RMSD-tolerance of 1.0 Å), of which 5 contained more than one member. The variation in energy from the best to the worst solution was 9.3 kcal/mol. There was one cluster of two solutions in the top 13, but the largest cluster began with the $14^{th}$-ranked solution. Within 1.0 Å R.M.S.D., there are 5 similar solutions. In total, 11 solutions have the B face of the trigonal protofilament model at the light-chain CDR, with slight overlap onto heavy chain CDR. One of them is offset by one monomer, others by acute angles. This is the largest supercluster; its lowest-energy solution is shown in **Figure 5b**. This solution resembles the supercluster solution found in the WO1 docking, but makes a 16° angle to it. The particular solution used for the following analysis was the lowest-energy member of this supercluster, which is 6.4 kcal/mol less favorable than the lowest-energy solution.

**4.5.1 Salt Bridges and Ion Pairs**
The analysis for this docked conformation is very similar to that of the WO1-Aβ6 solution, but there are no salt bridges at less than 4.0 Å, only two ion pairs between 4.0 and 6.0 Å, and five between 6.0 and 8.0 Å. *Hydrogen Bonding and hydrophobic interactions.* Thirteen likely hydrogen-bonds at less than 3.5 Å exist between WO2's Fv model and the docked trigonal-prism model. No hydrophobic interactions were observed for this docking model.

**4.6 Docking. WO1-Two-Strand-Model**
AutoDock3 predicted 50 binding solutions. Both the Fv and the protofilament model were treated as rigid bodies. For WO1 and the two-strand protofilament model, Autodock3 found 4 distinct conformational clusters with more than two members (using an RMSD-tolerance of 2.0 Å. The variation in energy from the best to the worst solution was 5.6 kcal/mol. In total, 4 solutions have the N-terminal face on the CDR, 4 have the open end of the two-strand model sitting longitudinally on the CDR, 7 solutions have the "bottom" end (right-hand-rule) of the two-strand protofilament model laterally across the CDR, and 3 have the closed end of the two-strand model near the CDR. The several clusters of near-equal energy suggest a negative result; they are shown in **Figure 6a**. Two of these clusters are not predicted to make any salt bridges at < 4 Å, but the other two are expected to make one and two such bonds, respectively.

**4.7 Docking. WO2-Two-Strand-Model**
As before, AutoDock3 predicted 50 binding solutions with both components (the Fv and the protofilament model) treated as rigid bodies. For WO2 and the two-strand protofilament model, Autodock3 found 35 distinct conformational clusters (using an RMSD-tolerance of 2.0 Å), of which 3 contained more than two members. The variation in energy from the best to the worst solution was 8.6 kcal/mol. The top 7 solutions have the "bottom" end (right-hand-rule) of the two-strand protofilament model laterally across the CDR. This cluster averages 3.8 kcal/mol lower in energy than the next best solution and it is the largest cluster, indicating a positive result; it is displayed in **Figure 6b**. This somewhat resembles the 7-member cluster for WO1-two-strand-model (shown in **Figure 6a**). The particular solution used for the following analysis was the lowest-energy member of the cluster.

**4.7.1 Salt Bridges, Ion Pairs, Hydrogen Bonding, and Hydrophobic Interactions**
No salt bridges were observed out to 4.0 Å, but 10 ion pairs in the range of 4-6 Å suggest a modest electrostatic contribution to WO2 binding Aβ. WO2 and the two-strand model make 7 likely hydrogen bonds at less than 3.4 Å. No *pi*-overlaps are observed.

**4.8 Docking Controls**
Positive controls #1 and #2 used rigid models, positive control #3 allowed flexibility in the ligand. Positive control #1 successfully recreated the crystal structure 1DQJ,[66] lysozyme complexed to antilysozyme. Two of the 50 solutions clustered near that from the crystal structure. These two averaged 11.8 kcal/mol lower in energy than the next most energetically favorable solution. Positive control #2, mouse renin complexed to the angiotensin analog CH-66 (PDB code 1SMR),[67] showed a similar result, with the 44-member dominant cluster 19.7 kcal/mol lower in energy. Positive control #3 did not precisely recreate the crystal structure 1DYI,[68] DHFR complexed with

folate, but rather predicted 10 similar solutions for the 32-atom folate molecule in the correct binding pocket within 2.3 to 10.8 Å RMSD of the true crystal structure, and within 3.5 kcal/mol of one another. Negative control #1 consisted of "docking" the trigonal prism protofilament model with the crystal structure of mouse renin. The result consisted of three superclusters (23, 14, and 5 members) of solutions whose lowest-energy members were within 5.0 kcal/mol of each other. Negative control #2 consisted of "docking" the WO1 Fv model with the inhibitor from the mouse renin crystal structure. All fifty top solutions were within 4 kcal/mol of one another. There was one two-member cluster and one four-member cluster. Solutions ranged all over the Fv.

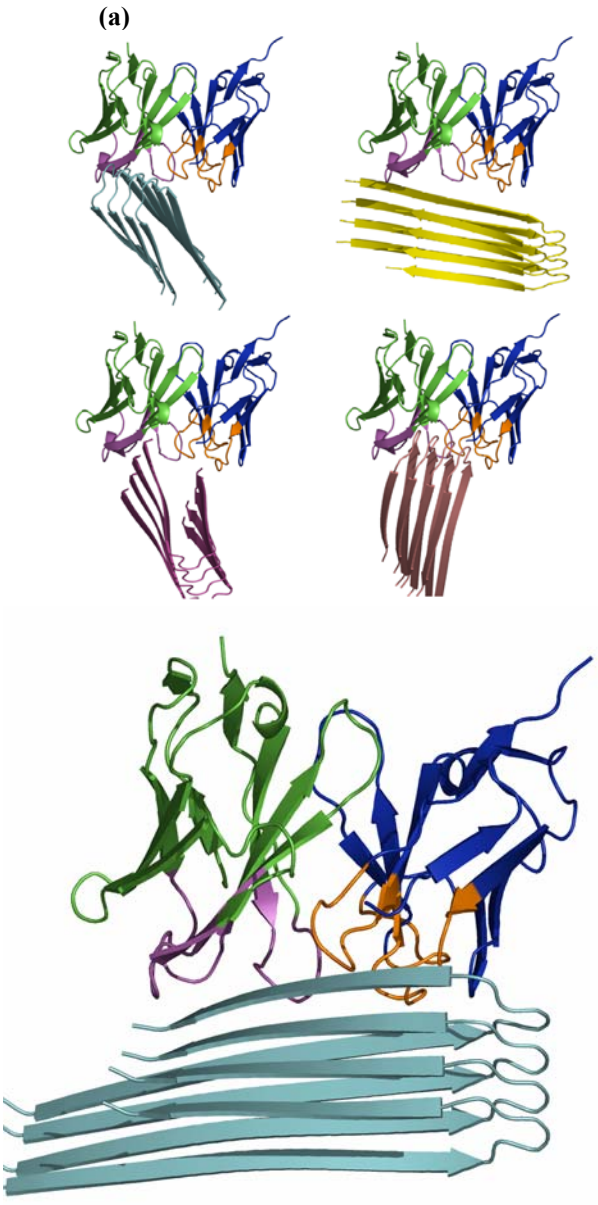



**Figure 6.** **(a)** WO1 Fv docked with two-strand protofilament model in the four predicted configurations. **(b)** WO2 Fv docked with two-strand protofilament model in the top predicted configuration

### 4.9 Binding vs. Salt Concentration

Binding constants were calculated from the sigmoid midpoints of the graph in **Figure 7**. The strongest binding of WO1 and Aβ was observed in PBS; the $EC_{50}$ was 2 nM. Of the 5 mM Hepes conditions, the ones at 150 mM NaCl (nearly-physiological) and at no salt resulted in the strongest binding of WO1 and Aβ, 4 nM. In the presence of 600

mM NaCl, the $EC_{50}$ constant changed sharply to 20 nM, indicating five-fold less binding. Moreover, the 150 mM KCl and 75 mM $Na_xH_yPO_4$ conditions resulted in 8 nM and 10 nM EC50's, respectively. The significantly reduced binding in high salt suggests that electrostatic interactions are relatively important to binding and hydrophobic interactions are relatively unimportant.

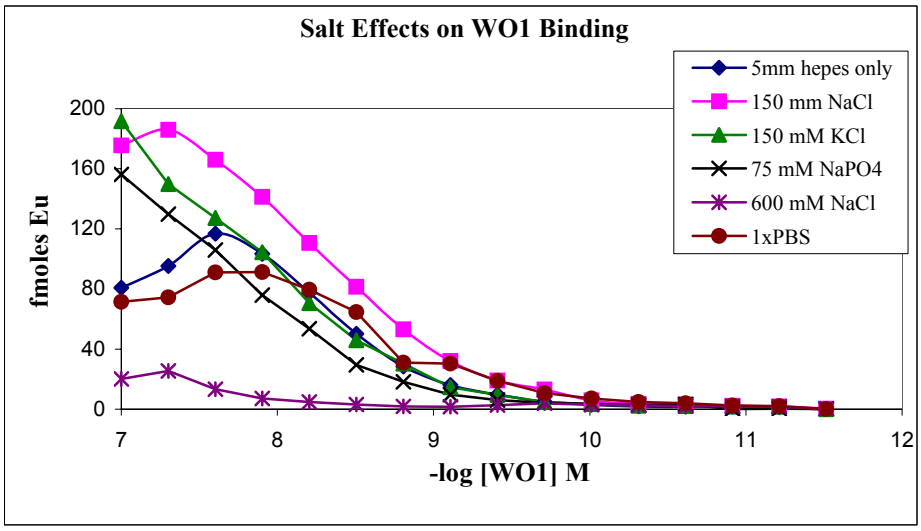



**Figure 7. Salt effect on WO1 binding.** Binding under various salt conditions was calculated from sigmoid midpoints. The $EC_{50}$ values are given in the text.

WO1 binding to Aβ was tested in a range of pH conditions. The Citrate pH 3.0 condition resulted in essentially no binding. Citrate and PIPES at pH 5.8 showed $EC_{50}$ values of 2 nM and >20 nM, respectively. The $EC_{50}$ values for PIPES and Bicine at pH 7.4 were 2 nM and 1.5 nM. For Bicine and CHES at pH 8.5, $EC_{50}$ values were 16 nM and 10 nM. For CHES at pH 10.0, the $EC_{50}$ was > 50 nM. (**Figure 8**). The data suggest that WO1 binds Aβ amyloid best closer to physiological pH values. If hydrophobic interactions dominated WO1- Aβ amyloid binding, this would likely not be the case. This supports the salt effect results suggesting that binding depends at least in part on electrostatic interactions.

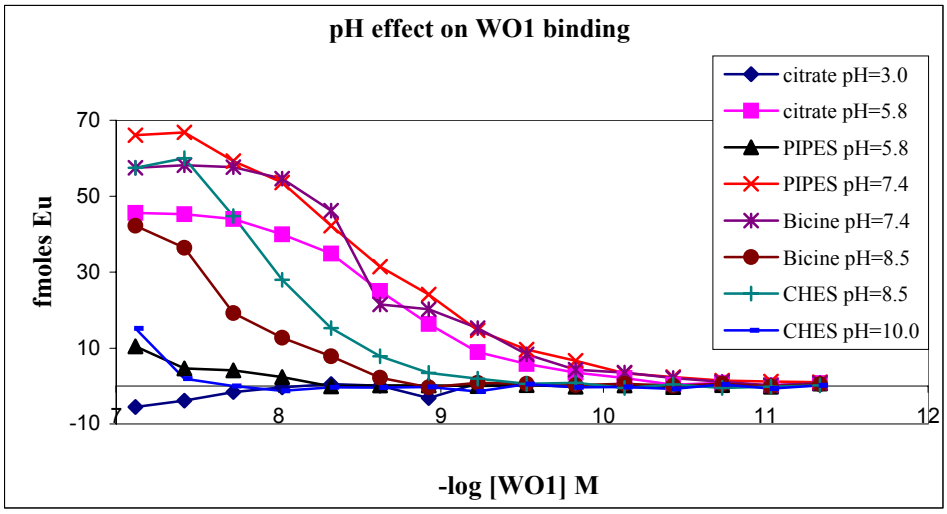



**Figure 8. pH effect on WO1 binding.** The show binding under various pH conditions. Binding affinities were calculated from the sigmoid midpoints. The $EC_{50}$ values are given in the text.

## 5. DISCUSSION

### 5.1 Comparison with Experimental Results

The model of WO1- trigonal prism-model docking that we have proposed, shown in **Figure 5a**, features many hydrogen-bonding pairs and salt-bridges; that agrees with the experimental salt-effect data for WO1- Aβ binding. However, the salt-binding data itself is at odds with WO1's general amyloid-recognition behavior which includes binding to (uncharged) polyglutamine fibrils. This prediction is also called into question by recent data from hydrogen exchange NMR[39] and cysteine scanning mutagenesis[40]. These data are not straightforward to interpret, but they suggest that residues 23-27 (the so-called "B face" of the trigonal-prism protofilament model) may not be in ordered β-sheet structure. There are no strong salt bridges at less than 4.0 Å predicted for binding of WO2 to either the trigonal prism model or the two-strand model.

### 5.2 Assemblies

The Fv models presented here are docked only to a protofilament not to the full Aβ fibril, but these docked models are consistent with the hierarchical fibril model for Aβ amyloid proposed by their authors.[26,38] The fibril model based on the trigonal prism protofilament construction consists of six trigonal prisms packed vertically against one another in two rows of three, in which each protofilament face is exposed twice on the outside surface of the fibril model. For the two-strand model, a full protofilament is generated by applying a two-fold symmetry axis along the C-terminal face perpendicular to the stacking axis of the model. In this way, the open ends of both stacks of monomers point the same way as the stacks twist helically around one another. One depiction of this twisting stack may be found in this work, **Figure 1e**; another is in Petkova, *et al* PNAS 2002, figure 5a. This implies that the C-terminal face of the model is not available for binding; as the predicted docking mode does not employ the C-terminal face, this does not affect our analysis. The Fv model likewise has a single antigen-recognition region for WO1/WO2, whereas the physiological IgM presents ten recognition sites, resulting in high avidity and increased total binding strength.

### 5.3 Implications of Trigonal Prism Results

The docking results for the trigonal prism model suggest that WO1 and WO2 bind to the face of amyloid fibrils similarly. The in-register parallel beta-stacking of the trigonal prism model requires that each residue of the polypeptide line up with its equivalent in the next layer of the prism. Thus, while Glu22 makes peptide bonds to Ala21 and Asp23, it makes hydrogen bonds to Glu22 on neighboring strands. Viewing a face of the prism, then, one would see lines of equivalent residues. The line of Glu22 and Asp23 residues (**Figure 4**) is predicted to bind to the corresponding line of positive and polar residues of WO1's light-chain CDR. The line of Asn27 and Lys28 residues likewise interact with the trail of negative and polar residues.

The lineup of positive and negative charges on the B face with those at the CDR suggests another available binding mode. The A face of the model also has such a lineup of charges (Figure 3). The N-terminal residues of the trigonal prism model occlude part of the A face in this model, which was not allowed torsional freedom, so the A face was less accessible for binding than the B face. Even so, 2/50 solutions for WO1 and 5/50 solutions for WO2, though not clustered, did predict some form of docking to the A face. It is worth noting, furthermore, that earlier simulations with an un-minimized model (in which the N terminal residues took other orientations) yielded a supercluster of docking solutions at the A face of the trigonal prism model with WO1. It is thus likely that a more computationally intensive docking simulation, one which allowed rotation around the bonds of the trigonal prism model, would have shown a bimodal distribution of solutions, one at the A face and another at the B face. These facts, coupled with the new data that suggest non β-sheet conformation for the B-face,[39,40] to some degree weaken the case for the trigonal-prism model for Aβ protofilament. Therefore WO1/WO2 binding to the A face of Aβ remains a distinct possibility. As only 4/50 solutions featured the C face, and they all took very different orientations, our current docking model suggests that the C face is unlikely to feature in WO1 binding.

### 5.4 Implications of Two-Strand Results

Docking calculations based on the two-strand model do not indicate a conclusive result for WO1-binding. However, computations for WO2 support WO2 binding to the end of elongating amyloid fibrils, while only the 48$^{th}$ ranked hit for WO2 and the trigonal prism resembles such an arrangement. This contrasts with expectation that the lines of charge along the N-terminal face would provide a recognition site for the antibodies, but is consistent with the fact that these antibodies bind amyloid fibrils (such as polyglutamine) that lack charged residues. Also, the lower degree

of order found in residues 23-27 in the two-strand model may be more consistent with the new hydrogen exchange and cysteine mutagenesis data.

## 6. CONCLUSIONS

The sequences of two amyloid-recognizing IgMs were compared and analyzed. Structural models of their Fvs were constructed. WO1's Fv was found to be unusually rich in charged residues, while WO2 is rich in Ser residues. As one might expect from the differences in their secondary and ternary structures, the two different protofilament models lead to different predicted docking models. Results of docking simulation for the trigonal-prism model imply that WO1 and WO2 bind to the face of a fibril of Aβ. The results for the two-strand model make no positive prediction for WO1, but results for the two-strand model suggest that WO2 binds to the ends of fibrils. The results for WO1 and the trigonal prism model agree with experimental salt-binding data, while the results for WO2 and the two-strand model are more consistent with WO1/WO2's diverse binding behavior. While these results do not support either model over the other, they make testable predictions.

## 7. ACKNOWLEDGMENTS

We thank Doctors Juntao Guo and Ying Xu for the trigonal prism protofilament model and Doctor Robert Tycko for the two-strand protofilament model.

## 8. REFERENCES

[1] R. H. Falk, R. L. Comenzo and M. Skinner, *N Engl J Med* 337 (1997).
[2] D. J. Selkoe, *Ann N Y Acad Sci* 924 (2000), 17.
[3] I. Tariska, *Acta Med Acad Sci Hung* 21 (1965), 483.
[4] P. A. Temussi, L. Masino and A. Pastore, *Embo J* 22 (2003), 355.
[5] E. T. Jaikaran and A. Clark, *Biochim Biophys Acta* 1537 (2001), 179.
[6] M. R. Hayden and S. C. Tyagi, *Jop* 2 (2001), 124.
[7] O. Corti and A. Brice, *J Soc Biol* 196 (2002), 95.
[8] R. Kruger, O. Eberhardt, O. Riess and J. B. Schulz, *Trends Mol Med* 8 (2002), 236.
[9] J. M. George, *Genome Biol* 3 (2002), REVIEWS3002.
[10] M. Pras, F. Prelli, E. C. Franklin and B. Frangione, *Proc Natl Acad Sci U S A* 80 (1983), 539.
[11] H. D. Feiner, *Hum Pathol* 19 (1988), 1255.
[12] J. N. Buxbaum, J. V. Chuba, G. C. Hellman, A. Solomon and G. R. Gallo, *Ann Intern Med* 112 (1990), 455.
[13] V. Bellotti, P. Mangione and G. Merlini, *J Struct Biol* 130 (2000), 280.
[14] C. A. Ross, *Neuron* 35 (2002), 819.
[15] C. Weissmann, *Nat Rev Microbiol* 2 (2004), 861.
[16] H. Inouye, P. E. Fraser and D. A. Kirschner, *Biophys J* 64 (1993), 502.
[17] M. Sunde and C. Blake, *Adv Protein Chem* 50 (1997), 123.
[18] S. B. Malinchik, H. Inouye, K. E. Szumowski and D. A. Kirschner, *Biophys J* 74 (1998), 537.
[19] A. D. Williams, et al., *J Mol Biol* 335 (2004), 833.
[20] J. D. Harper, S. S. Wong, C. M. Lieber and P. T. Lansbury, Jr., *Biochemistry* 38 (1999), 8972.
[21] L. C. Serpell, et al., *J Mol Biol* 300 (2000), 1033.
[22] T. C. Saido, W. Yamao-Harigaya, T. Iwatsubo and S. Kawashima, *Neurosci Lett* 215 (1996), 173.
[23] I. Kheterpal, A. Williams, C. Murphy, B. Bledsoe and R. Wetzel, *Biochemistry* 40 (2001), 11757.
[24] O. N. Antzutkin, R. D. Leapman, J. J. Balbach and R. Tycko, *Biochemistry* 41 (2002), 15436.
[25] M. Torok, et al., *J Biol Chem* 277 (2002), 40810.
[26] A. T. Petkova, et al., *Proc Natl Acad Sci U S A* 99 (2002), 16742.
[27] T. S. Burkoth, et al., *J. Am. Chem. Soc.* 122 (2000), 7883.
[28] I. Kheterpal, et al., *Biochemistry* 42 (2003), 14092.
[29] S. S. Wang, S. A. Tobler, T. A. Good and E. J. Fernandez, *Biochemistry* 42 (2003), 9507.
[30] M. Kraus, M. Bienert and E. Krause, *Rapid Commun Mass Spectrom* 17 (2003), 222.
[31] I. Kheterpal, S. Zhou, K. D. Cook and R. Wetzel, *Proc Natl Acad Sci U S A* 97 (2000), 13597.
[32] T. L. Benzinger, et al., *Proc Natl Acad Sci U S A* 95 (1998), 13407.
[33] M. F. Perutz, J. T. Finch, J. Berriman and A. Lesk, *Proc Natl Acad Sci U S A* 99 (2002), 5591.
[34] L. O. Tjernberg, et al., *Biochem J* 366 (2002), 343.

[35] C. Hilbich, B. Kisters-Woike, J. Reed, C. L. Masters and K. Beyreuther, *J Mol Biol* 218 (1991), 149.
[36] M. O. Chaney, S. D. Webster, Y. M. Kuo and A. E. Roher, *Protein Eng* 11 (1998), 761.
[37] L. C. Serpell, C. C. Blake and P. E. Fraser, *Biochemistry* 39 (2000), 13269.
[38] J. T. Guo, R. Wetzel and Y. Xu, *Proteins* 57 (2004), 357.
[39] N. A. Whittemore, et al., *Biochemistry* 44 (2005), 4434.
[40] S. Shivaprasad and R. Wetzel, *Biochemistry* 43 (2004), 15310.
[41] R. Nelson, et al., *Nature* 435 (2005), 773.
[42] B. O'Nuallain and R. Wetzel, *Proc Natl Acad Sci U S A* 99 (2002), 1485.
[43] M. Dumoulin and C. M. Dobson, *Biochimie* 86 (2004), 589.
[44] A. C. Davis and M. J. Shulman, *Immunol Today* 10 (1989), 118.
[45] N. R. Whitelegg and A. R. Rees, *Protein Eng* 13 (2000), 819.
[46] Chemical Computing Group Inc., 1997.
[47] E. E. Abola, J. L. Sussman, J. Prilusky and N. O. Manning, *Methods Enzymol* 277 (1997), 556.
[48] J. L. Sussman, et al., *Acta Crystallogr D Biol Crystallogr* 54 (1998), 1078.
[49] A. C. Martin, *Proteins* 25 (1996), 130.
[50] S. F. Altschul, et al., *Nucleic Acids Res* 25 (1997), 3389.
[51] C. Chothia and A. M. Lesk, *J Mol Biol* 196 (1987), 901.
[52] T. T. Wu and E. A. Kabat, *Mol Immunol* 29 (1992), 1141.
[53] A. C. Martin, J. C. Cheetham and A. R. Rees, *Proc Natl Acad Sci U S A* 86 (1989), 9268.
[54] A. C. Martin, J. C. Cheetham and A. R. Rees, *Methods Enzymol* 203 (1991), 121.
[55] A. R. Rees, Searle,S.M.J., Henry,A.H.and Pedersen,J.T., *in* "Protein Structure Prediction" (S. M.J.E., ed.), p. 141–172. Oxford University Press, Oxford, 1996.
[56] J. T. Pedersen, Searle,S.M.J., Henry,A.H. and Rees,A.R., *Immunomethods* 1 (1992), 126.
[57] G. M. Morris, et al., *Journal of Computational Chemistry* 19 (1998), 1639.
[58] S. J. K. Weiner, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P., *J Am Chem Soc* 106 (1984), 765.
[59] S. J. K. Weiner, P. A.; Nguyen, D. T.; Case, D. A., *J Comput Chem* 7 (1986), 230
[60] Accelrys; MSI, San Diego, 2000.
[61] G. Murshudov, *Acta Crystallogr D Biol Crystallogr.* 53 (1997), 240.
[62] N. Collaborative Computational Project, *Acta Crystallogr D* 50 (1994), 760.
[63] C. Cole and J. Warwicker, *Protein Sci* 11 (2002), 2860.
[64] W. L. DeLano. DeLano Scientific, San Carlos, CA, USA, 2002.
[65] A. Szilagyi and P. Zavodszky, *Structure Fold Des* 8 (2000), 493.
[66] Y. Li, H. Li, S. J. Smith-Gill and R. A. Mariuzza, *Biochemistry* 39 (2000), 6296.
[67] C. G. Dealwis, et al., *J Mol Biol* 236 (1994), 342.
[68] V. M. Reyes, M. R. Sawaya, K. A. Brown and J. Kraut, *Biochemistry* 34 (1995), 2710.

# SNAPSHOTS OF THE PANTOTHENATE SYNTHETASE FROM MYCOBACTERIUM TUBERCULOSIS ALONG THE REACTION COORDINATE

Shuishu Wang
*Public Health Research Institute, 225 Warren Street, Newark, New Jersey 07103*

David Eisenberg
*Howard Hughes Medical Institute, UCLA-DOE Institute of Genomics and Proteomics, Molecular Biology Institute, UCLA, Box 951570, Los Angeles, CA 90095-1570*

## 1. ABSTRACT

Pantothenate synthetase (PS) from *Mycobacterium tuberculosis* represents a potential target for developing new anti-tuberculosis drugs. PS catalyzes the ATP-dependent condensation of pantoate and β-alanine to form pantothenate. We determined the crystal structures of PS from *M. tuberculosis* and its complexes with substrates ATP, pantoate, and β-alanine, as well as an ATP analog AMPCPP, a reaction intermediate pantoyl adenylate and a reaction product AMP, with resolutions from 1.6 to 2 Å. The PS structure reveals a dimer, and each subunit has two domains with tight association between domains. The active site cavity is on the N-terminal domain, covered by the C-terminal domain. There is a flexible loop that forms one wall of the active site cavity, and it opens and closes the active site cavity. The enzyme binds ATP and pantoate tightly in the active site, and brings the carboxyl oxygen of pantoate near the α-phosphorus atom of ATP for an in-line nucleophilic attack. The pantoyl adenylate thus formed is stabilized by tight binding interactions with enzyme active site residues, and protected from hydrolysis by closing of the active site cavity with the flexible loop becoming ordered. Binding of β-alanine can occur only after formation of the pantoyl adenylate intermediate. Crystal structures of these complexes provide a step-by-step view of the PS catalyzed reaction, and allow us to construct transient reaction intermediates and deduce detailed reaction mechanism. PS catalyzes the reaction by stabilizing the reaction intermediates along the reaction coordinate, and structural models of these intermediates can provide basis for inhibitor design.

## 2. INTRODUCTION

Pantothenate (vitamin B5) is an essential precursor for the biosynthesis of coenzyme A and acyl carrier proteins, both of which play critical roles in many cellular processes including energy metabolism and fatty acid metabolism [1]. Microorganisms and plants can synthesize pantothenate, while animals obtain this essential nutrient from their diet [2]. Therefore, the pantothenate biosynthetic pathway offers targets for developing drugs against microbial pathogens. Recently, Jacobs and coauthors [3] reported that a *Mycobacterium tuberculosis* (MTB) mutant defective in the de novo biosynthesis of pantothenate is highly attenuated in both immunocompromised and immunocompetent mice. This observation indicates that a functional pantothenate biosynthetic pathway is essential for virulence of MTB, and thus represents a new target for anti-tuberculosis agents.

The pantothenate biosynthetic pathway in bacteria comprises four steps catalyzed by enzymes encoded by the *panB*, *panC*, *panD* and *panE* genes [4]. The *panC* gene encodes a pantothenate synthetase (PS), which catalyzes the last step of pantothenate biosynthesis, the ATP-dependent condensation of pantoate and β-alanine to form pantothenate. The *panC* gene product in several organisms has been identified and characterized, including those in *Escherichia coli* [5], *Saccharomyces cerevisiae*, higher plants *Oryza sativa* and *Lotus japonicus* [6], fungus *Fusarium oxysporum* [7], and *M. tuberculosis* [8]. The PS enzymes from *E. coli*, higher plants, and MTB form dimers in solution.

The *E. coli* PS enzyme structure belongs to the cytidylyltransferase superfamily [5]. It has two distinct domains, a large N-terminal domain having a Rossmann fold and a smaller C-terminal domain containing a helical layer above a three-stranded antiparallel β-sheet. Based on structural comparison of the *E. coli* PS with other members of cytidylyltransferase superfamily having known structures, von Delft et al. [5] deduced the ATP and pantoate binding sites of the *E. coli* PS and proposed a hinged domain mechanism for opening and closing of the enzyme active site cavity.

Kinetic analysis of the MTB PS suggests that the enzyme-catalyzed reaction proceeds through two steps: the formation of an enzyme bound intermediate, pantoyl adenylate, from ATP and pantoate, followed by nucleophilic attack on the intermediate by β-alanine to form pantothenate and AMP [8]. The existence of pantoyl adenylate as an enzyme bound intermediate is suggested by the transfer of the $^{18}O$ label from the carboxyl group of pantoate to the reaction product AMP. However, pantoyl adenylate cannot be directly isolated or identified by TLC methods, presumably due to rapid lactonization [9].

Here, we present the crystal structure of the MTB PS enzyme and structures of its complexes with substrates, a substrate analog (AMPCPP), a reaction product (AMP), and a pantoyl adenylate reaction intermediate [10, 11]. The MTB PS has the same fold as the *E. coli* enzyme. However, in the MTB PS structure the domains of each subunit have a closed conformation, in contrast to those of the *E. coli* PS structure [5], and there is no significant movement between domains among all structures of the apo-enzyme and various complexes. A flexible region, which forms a wall of the active site cavity, becomes ordered in the reaction intermediate complex and closes the active site cavity, thus acting as a gate to the active site cavity of the MTB PS enzyme. The structure of the β-alanine complex gives direct evidence that the binding site for this last substrate exists only after formation of the pantoyl adenylate intermediate. Together these crystal structures lead to a detailed view of the overall reaction mechanism. Models of the transient reaction intermediate can be easily constructed from the experimentally observed complexes.

## 3. MATERIALS AND METHODS

### 3.1 Protein Production

The detailed procedures for cloning of the *panC* gene, expression and purification of the protein have been described [10]. Briefly, the MTB *panC* gene (Rv3602c) encoding the pantothenate synthetase was amplified from the genomic DNA of MTB strain H37Rv and inserted into a pET30a plasmid (Novagen), which produces a recombinant protein with an N-terminal 6xHis tag that can be cleaved off with enterokinase. BL21(DE3) cells containing the pET30-*panC* plasmid were grown in LB medium containing 50 μg/ml kanamycin at 37 °C and induced with 0.4 mM IPTG for 3 hours. Cells were collected and lysed by lysozyme treatment followed by sonication in 20 mM HEPES pH 7.8, 500 mM NaCl, and 0.5 mM PMSF. The protein was purified from a $Ni^{2+}$-charged HiTrap chelating column (Amersham) with 20 mM HEPES pH 7.8, 500 mM NaCl and a linear gradient of imidazole. The PS protein was subjected to enterokinase (New England Biolabs) digestion to cleave off the N-terminal fusion tag. Electrospray mass spectroscopy indicated that 9 residues from the C-terminus of the protein were also cleaved off by enterokinase digestion. However, enzyme activity assays suggested that this C-terminal truncation does not have any detectable effect on the enzyme activity [10]. The enterokinase digested protein was further purified with a second round of chromatography on a $Ni^{2+}$-HiTrap column followed by Superdex 75 column (Amersham) gel filtration.

### 3.2 Crystallization and Data Collection

Crystallization was carried out as described previously [10]. Crystals in space group $P2_1$ were obtained from drops set up with well solutions containing 10-15% PEG 3000, 5% glycerol, 2% ethanol, 20 mM $MgCl_2$, 150 mM $Li_2SO_4$ and 100 mM imidazole pH 8.0 at 20 ºC. Those in space group C2 were from similar conditions, except with 2% isopropanol in place of ethanol and with 200 mM $Li_2SO_4$. Cocrystallization with β-alanine was done with a well solution of 15% PEG 3000, 2% ethanol, 20 mM $MgCl_2$, 100 mM imidazole pH 8.0 and 20 mM β-alanine, from which crystals in space group $P2_12_12_1$ were obtained. However, these crystals do not have β-alanine in the active site. Cocrystallization with both ATP and pantoate was carried out in the same crystallization condition of the $P2_1$ crystals but with both ATP and pantoate added, from which isomorphous $P2_1$ crystals were obtained that have a pantoyl adenylate intermediate in the active site.

Crystals in complex with substrates were obtained by soaking apo-enzyme crystals in solutions containing substrates or the ATP analog AMPCPP. Soaking experiments were carried out by adding solutions containing substrates or AMPCPP directly to the drops, or to drops of pseudo mother liquor with transferred crystals, and incubating overnight or longer. Crystals were soaked in solutions containing ~10 mM each of AMPCPP alone, both AMPCPP and pantoate, both pantoate and β-alanine, both ATP and pantoate, and both AMP and β-alanine. Crystals were also soaked in solutions containing up to 20 mM pantothenate, but no pantothenate was found in the crystal structure.

Before data collection, crystals were soaked for 2-5 min in a cryogenic solution similar to well solutions or substrate soaking solutions with glycerol added to 30%, and they were then flash-frozen in a cryo stream of $N_2$ gas at 100 K. Diffraction data were collected at 100 K on a Rigaku FRD generator with an R-AXIS $IV^{++}$ detector. Data reduction and scaling were carried out with the programs DENZO and SCALEPACK [12]. Data processing statistics are reported in references [10] and [11]. Table 1 summarizes a few of the data sets, of which the structures are mentioned in this paper. The $P2_1$ crystal has two PS molecules per asymmetric unit; while the C2 crystal contains one molecule per asymmetric unit. In the C2 crystal, the active site is locked in an inactive conformation by the crystal packing (see below), and the pantoate and β-alanine binding sites are compromised. Therefore, only AMP was found when we soaked the C2 crystals with both AMP and β-alanine (data set 8, in Table 1), and those soaked with both ATP and pantoate gave an ATP complex (data set 9) instead of a reaction intermediate complex.

**TABLE 1**

Summary of crystal structure data and experiments

| | Complex | Experiment | SG | #/A.U. | resolution | PDB ID | Ref. |
|---|---|---|---|---|---|---|---|
| 1 | Apo enzyme | | $P2_1$ | 2 | 1.6 Å | 1MOP | [10] |
| 2 | AMPCPP and pantoate in A; AMPCPP in B | AMPCPP, pantoate, soak | $P2_1$ | 2 | 1.6 Å | 1N2E | [10] |
| 3 | Pantoyl adenylate | ATP, pantoate, cocrystal | $P2_1$ | 2 | 2.0 Å | 1N2H | [10] |
| 4 | Pantoyl adenylate | ATP, pantoate, soak | $P2_1$ | 2 | 1.7 Å | 1N2I | [10] |
| 5 | Pantoate | Pantoate, β-alanine, soak | $P2_1$ | 2 | 1.8 Å | 1N2J | [10] |
| 6 | AMP, β-alanine in A; AMP only in B | AMP, β-alanine soak | $P2_1$ | 2 | 1.85 Å | 1A86 | [11] |
| 7 | Apo enzyme | | C2 | 1 | 1.7 Å | 2A88 | [11] |
| 8 | AMP | AMP, β-alanine soak | C2 | 1 | 1.7 Å | 2A7X | [11] |
| 9 | ATP | ATP, pantoate, soak | C2 | 1 | 1.6 Å | 2A84 | [11] |

### 3.3 Structural Determination and Refinement

The crystal structure of the apo-enzyme in the $P2_1$ crystal was determined [10] by the molecular replacement method using AMORE [13], with subunit A of the *E. coli* PS structure (PDB ID code 1IHO [5]) as a model. The two domains were separated as two independent search models. The positions of the two N-terminal domains and one C-terminal domain were found from the rotation and translation searches, and the other C-terminal domain was generated by a non-crystallographic-symmetry (NCS) operation on the position of the known C-terminal domain with the NCS symmetry operators derived from the two N-terminal domains. The structure of the apo-enzyme was used to calculate phases for data sets of various complexes by rigid body refinement into new data sets for isomorphous crystals, or by molecular replacement with AMORE. The final refinement statistics are reported in references [10, 11]. Structural models were refined against diffraction data using CNS [14] and/or REFMAC [15]. After each cycle of refinement, models were manually adjusted with electron density maps using O [16]. The same subsets of data for $R_{free}$ calculation were kept between the CNS and REFMAC programs.

Besides the 9 residues that were digested off by enterokinase, there are 10 more residues from the C-terminus that do not have clear electron density. Based on its susceptibility to protease digestion and lack of clear electron density, it is reasonable to assume that these residues at the C-terminus of the protein are flexible, and do not have a defined structure in solution. It is clear from the structure that both N- and C-termini are away from the active site cavity, and therefore are unlikely to affect the enzyme catalytic activity. This is consistent with the results of enzyme activity assays [10].

## 4. STRUCTURE OF THE MTB PANTOTHENATE SYNTHETASE

### 4.1 Dimer Structure

The crystal structure of the MTB PS enzyme indicates that it is a dimer (**Figure 1**), resembling a butterfly when viewed from one side. The dimer interface is extensive, having a buried area of about 2150 $Å^2$. A cluster of 8 hydrophobic side chains at the center of the dimer interface, Phe174, Leu177, Leu144, and Val118, form the core of the dimer interaction. This cluster is surrounded by hydrogen bonds and salt bridges, as well as the specific β strand

interactions of the two-stranded intersubunit β sheet that is composed of the strand β5 from each subunit. The MTB PS enzyme also exists in solution as a dimer [8, 10], as do PS enzymes from higher plants [6, 7].

The two subunits of the dimer in the $P2_1$ crystal are similar to each other, with an rmsd of 0.9 for all Cα atoms except the disordered residues. Most of the deviation lies at residues 74 to 88, and residues 260 to 265, where the crystal packing environments are different. Residues 74 to 88 (β3-$3_{10}$3-α3'-α3) form one wall of the active site cavity (see **Figure 2** and the active site cavity section b 31 In subunit A these residues are ordered with residues 74 to 76 having high B factors, while in subunit B re- 75 to 80 (helix α3'and nearby residues) do not have electron density, and the rest of the residues up to residue 88 are partially disordered with high B factors. This flexible wall of the active site cavity becomes ordered in the complex with the pantoyl adenylate reaction intermediate, suggesting their function as a gate to the active site cavity (see below pantoyl adenylate binding interactions).

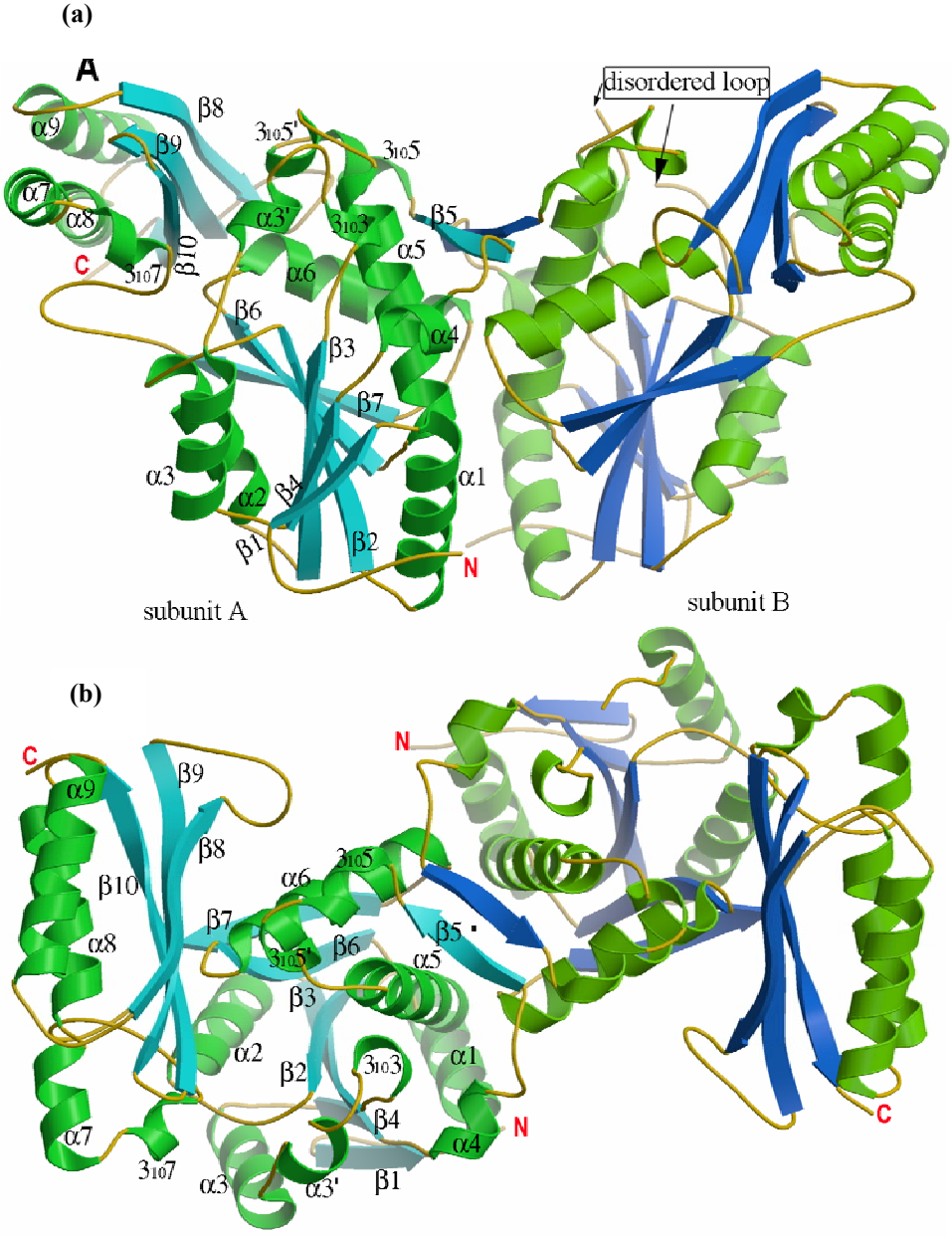



**Figure 1.** Ribbon diagram of the *M. tuberculosis* pantothenate synthetase dimer. **(a)** A side view of the dimer structure shows that it resembles the shape of a butterfly. **(b)** An orthogonal view of (A) from top, with the two-fold NCS symmetry axis (labeled with a dot) approximately perpendicular to the paper plane. Secondary structure elements for subunit A (left) are labeled. Those for subunit B are identical except that the short helix α3' is disordered. The figure was prepared from the apo-enzyme structure of the $P2_1$ crystal, with Molscript [17] and Raster3D [18].

The C2 crystal differs from the $P2_1$ crystal in that there is one polypeptide chain per asymmetric unit [11]. However, the protein molecule in the C2 crystal is also present as a dimer that is essentially identical to that in the $P2_1$ crystal, but the two-fold axis of the dimer coincides with one crystal symmetry axis. In the C2 crystal, a larger portion of the flexible wall, from residue 76 to 86, is completely disordered. The first turn of helix α3 unwinds, and helix $3_{10}3$, which contains residue Glu72 that is important for binding pantoate (see below pantoate binding site), also unwinds and forms an extended structure. This 32 ral difference is induced by crystal packing and the intrinsic flexibility of the loop.

### 4.2 Domain Interactions

Each subunit of the dimer has two well-defined domains (**Figure 1**). The N-terminal domain has a Rossmann fold, with a central parallel β sheet composed of strands β1 to β4, β6, and β7, and helices on both sides of the β-sheet. Strand β7 leads to the C-terminal domain, which starts with a hairpin loop followed by helix $3_{10}7$. The rest of the C-terminal domain has a simple two-layer structure: a layer of helices α7, α8, and α9 followed by a layer of antiparallel β sheet composed of strands β8 to β10.

The two domains have extensive interactions. There is a hydrophobic core composed of residues Leu123, Leu127, Tyr162, Val166 in the N-terminal domain and residues Tyr249, Leu257, and Leu269 in the C-terminal domain. Surrounding this hydrophobic core are three salt bridges, Glu159 to Arg267, Glu128 to Arg278, Glu126 to Arg253, and 6 additional hydrogen bonds linking the two domains. In addition, there are several water-mediated hydrogen bonding interactions. The relative positions of the two domains stay the same for the two subunits in the dimer and for all crystals including the $P2_1$, $P2_12_12_1$, and C2 crystals. These structural features suggest that domain hinge motion is unlikely to be the mechanism of opening and closing the active site cavity for MTB PS.

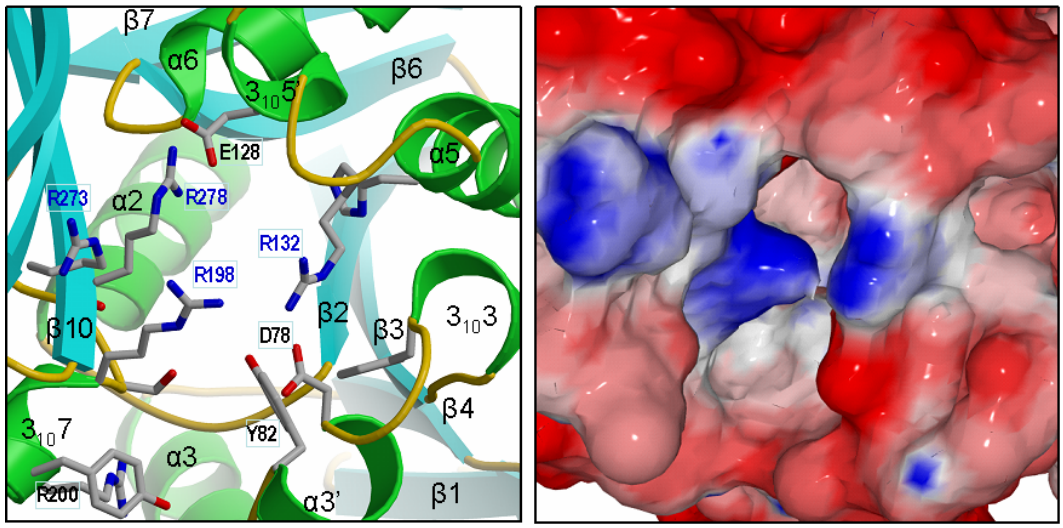



**Figure 2.** Active site cavity and the electrostatic potential surface around the entrance to the active site cavity. The figure was prepared from the subunit A of the $P2_1$ crystal. The two panels are approximately the same view. In the left panel, side chains at the entrance to the activity are shown as sticks. The segment from the end of β3 to the beginning of α3 forms the flexible wall, which is disordered in subunit B and in the C2 crystal thus opening the active site cavity.

### 4.3 Active Site Cavity

Typical for nucleotide binding (Rossmann fold) proteins, the active site is located at the C-terminus of the central parallel β sheet. The cavity is large and deep, and its bottom is at a cleft between strands β2 and β6 (**Figure 2**). The segment between β3 and α3 (loop-$3_{10}$3-loop-α3'-loop region) forms one side of the walls surrounding the cavity in subunit A of the $P2_1$ crystal. Part of this loop is disordered in subunit B, thus opens up the active site cavity. Other walls of the cavity are from β2-loop-α2, $3_{10}5'$-loop-α5, β6-loop-α6, and the loop after β7. Helix $3_{10}7$ and the β strands from the C-terminal domain partially cover the top of the active site cavity. The bottom of the active site cavity is mainly hydrophobic, while the top half of the cavity has several charged residues, including His44 and His47 at the N-terminus of α2 (the HIGH motif [5, 19]), Lys160 and Asp161 at the N-terminus of α6, and Arg198 of $3_{10}7$ (KMSKS motif [5]). At the entrance to the active site cavity, several charged side chains stick to the center,

leaving only a small opening to the cavity (**Figure 2**, right panel). Four arginine side chains, including Arg198, form a patch of positive charges at the entrance. These positive charges may steer the negatively charged substrates into the active site cavity. As we will see later, Arg198 plays an important role in binding the phosphate groups of ATP.

### 4.4 ATP Binding Interactions

The active site is well designed for binding ATP. **Figure 3** shows the initial difference electron density for the ATP molecule in the active site of the C2 crystal. The ATP molecule has an average temperature factor of 17.3 $Å^2$, virtually identical to the average temperature factor of the surrounding protein atoms, indicating that it is at full occupancy. A magnesium ion was found binding to the phosphate groups of ATP. It has a nearly perfect octahedral coordination. Its distances to the O2α, O2β, and O1γ atoms of the phosphate groups are 2.26 Å, 1.99 Å, and 2.12 Å, respectively; three other ligands are water molecules with distances of 1.96 Å, 2.11 Å, and 2.28 Å. The adenosine group of ATP binds tightly at the bottom of the active site cavity with hydrogen bonds and hydrophobic interactions [10, 11]. The phosphate groups stick toward the top of the active site cavity, and are located near the N-terminal ends of helices α2 and $3_{10}7$. Binding of the phosphate groups involves mainly main chain NH groups and charged side chains.

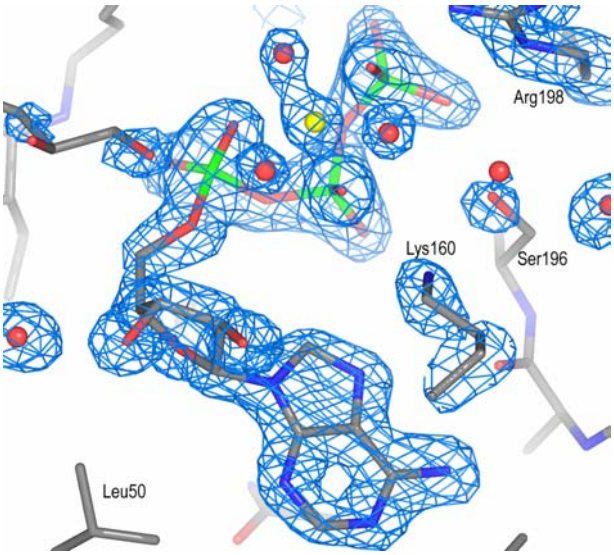



**Figure 3**. Initial Fo – Fc electron density in the active site of the C2 crystal soaked in a solution containing ATP and pantoate superimposed on the final refined model. The map was calculated at 1.6 Å and contoured at 3.5 σ. The ATP molecule has very well defined electron density. Side chains of Lys160, Ser196 and Arg198 moved relative to those in the apo-enzyme to interact with the phosphate groups, and thus have positive initial difference electron density. The figure is prepared with PYMOL [20].

Two His residues of the HIGH motif [5, 19, 21] interact with the phosphates of ATP. The first histidine, His44, has a hydrogen bond extending from its Nε2 to O1β of β-phosphate. The His47 side chain has a hydrogen bond from its Nε2 to the bridge oxygen between the α- and β-phosphate, suggesting that it functions as a general acid by donating a proton to the leaving pyrophosphate group. Both histidine residues have the Nδ1 atom forming a hydrogen bond to a main chain NH group, and thus the Nε2 atoms are hydrogen bond donors. Their imidazole rings are in close proximity to each other with the closest separation being ~3.2 Å, indicating favorable π-π interactions. These two side chains also play a role in binding the adenine group and stabilizing the reaction intermediate, because the imidazole rings are close to the adenine (closest distances of ~3.4 and ~4.3 Å for His44 and His47, respectively). Mutation of either histidine residue to alanine causes a greater than 1000-fold reduction in enzyme activity [22]. The Gly46 residue is highly conserved in this motif. It flanks one side of adenine (with Lys160 on the loop after β6 on the other side), where any side chain larger than a hydrogen atom would have steric hindrance. The KSMKS motif [23] is not conserved at the sequence level in PS enzymes, but it is evident from structural alignments [5]. Residues Ser196, Ser197, and Arg198, which are at the N-terminus of helix $3_{10}7$ and are part of the KSMKS motif, bind the β- and γ-phosphate groups. The side chains of Lys160 and Arg198 are partially disordered in the apo-enzyme structure, but become ordered and move to enable salt-bridges to the β- and γ-phosphate groups, respectively. The side chain of Lys160 also packs against one side of adenine. Mutation of this

residue to alanine reduces the enzyme activity more than 1000-fold, and it decreases affinity of the enzyme for ATP [22]. Another residue that binds ATP is Asp161, which has hydrogen bonds to two water ligands of the magnesium ion, and to O2* of the ribose group. Overall, the ATP molecule is rigidly held in the active site by many hydrophobic, π-electron, hydrogen bonding, and charge-charge interactions, which keep it in an optimal conformation for a nucleophilic attack by pantoate. Residues involved in binding the adenosine group are also important in stabilizing the pantoyl adenylate intermediate.

The binding interactions of ATP with protein are similar to those of AMPCPP [10]. However, the methylene group in AMPCPP replaces the bridge oxygen between the α- and β-phosphate groups of ATP, and thus it is not capable of forming a hydrogen bond with the Nε2 atom of His47. Consequently, the α-phosphate moves slightly away from amide nitrogen of Met40, and thus no hydrogen bond is formed between them in the AMPCPP complex. Therefore, ATP has slightly higher binding affinity to the enzyme active site than its non-reactive analog AMPCPP.

**4.5 Pantoate Binding Site**

In the $P2_1$ crystal that was soaked with pantoate and β-alanine, there is a full occupancy of pantoate in both active sites [10]. The binding site for pantoate is located in a pocket at the bottom of the active site cavity, next to the binding site of the adenosine group (**Figure 4a**). This is the same binding site proposed for the *E. coli* PS protein [5], although the pantoate molecule has a different conformation and therefore different binding interactions. The pantoate molecule is tightly bound at the bottom of the active site cavity, with side chains of two glutamine residues, Gln72 from helix $3_{10}3$ and Gln164 from helix α3, forming hydrogen bonds to its hydroxyl groups and one carboxyl oxygen. The two methyl groups face the hydrophobic groups at the bottom of the active site cavity, Pro38 and Phe157 side chains. The side chain of Met40 packs against the hydrophobic side of the pantoate molecule. Mutation of either glutamine residue to alanine greatly reduces the rate of pantoyl adenylate formation [22].

Soaking $P2_1$ crystals in solutions containing both AMPCPP and pantoate gave a full occupancy of AMPCPP, but with a glycerol in the pantoate binding site in subunit B, which has the flexible wall (residues 75-88) disordered. However, subunit A has a partial occupancy of both AMPCPP and pantoate, as indicated by their electron density and B factors. Increasing the concentration of pantoate in the soaking solution gave only a slightly higher occupancy of pantoate but a lower occupancy of AMPCPP [10]. All these observations indicate that AMPCPP and pantoate cannot coexist in the active site. This must be due to the steric and charge repulsions between the α-phosphate group and the carboxyl group of pantoate. Modeling both ATP and pantoate in one active site based on the fully occupied complexes will put one carboxyl oxygen atom of pantoate ~2.7 Å from the α-phosphorus atom of ATP [11]. This is essentially identical to what is observed in subunit A that has partial occupancy of both AMPCPP and pantoate. The carboxyl oxygen of pantoate is in a good position for an in-line nucleophilic attack on the α-phosphate. Because residues of the flexible wall (gate residues) are disordered in subunit B, the active site cavity is open and allows AMPCPP to diffuse easily into the active site. The extensive binding interactions of AMPCPP to the protein atoms make its binding thermodynamically more favorable than the binding of pantoate to the protein. Therefore, AMPCPP has a full occupancy in the active site cavity of subunit B of the complex with both AMPCPP and pantoate. On the other hand, the gate residues are less flexible in the subunit A due to crystal packing, thus making it kinetically less favorable for AMPCPP to bind, and allowing pantoate to have a partial occupancy in the active site of subunit A.

In the C2 crystal, the side chain of Gln72 on helix $3_{10}3$ swings out of the active site due to crystal packing, and the binding of pantoate is abolished. We observed only ATP in the active site with a full occupancy (see above ATP binding interactions) when we soaked the C2 crystal in a solution containing both ATP and pantoate [11]. There was no electron density for pantoate or pantoyl adenylate in the active site cavity. In the pantoate binding pocket, there was positive difference electron density, into which we modeled a glycerol molecule. This glycerol has a high temperature factor, suggesting that it has a low occupancy. Except for this Gln72 side chain, the rest of the binding site for pantoate is intact. In the $P2_1$ crystal, soaking with both ATP and pantoate results in a pantoyl adenylate intermediate in the active site cavity, and the disordered loop in subunit B becomes ordered [10] (see more details below). This indicates that loss of this glutamine side chain completely abolishes the enzyme catalytic activity for the first half of the reaction inside the crystal. The glutamine side chains are also involved in binding the reaction intermediate. Thus mutation of either residue could affect the stability of pantoyl adenylate intermediate. It is noteworthy that Asn69 was also found to be important for the enzyme activity [22]. This residue is not directly involved in binding substrates or the intermediate. However, it forms a hydrogen bond to the side chain of Gln72, and thus is important for keeping the Gln72 side chain in position. In addition, since both Asn69 and Gln72 are on

the one-turn $3_{10}3$ helix, mutation of Asn69 might also affect the stability of this $3_{10}$ helix and hence the position of the Gln72 side chain.

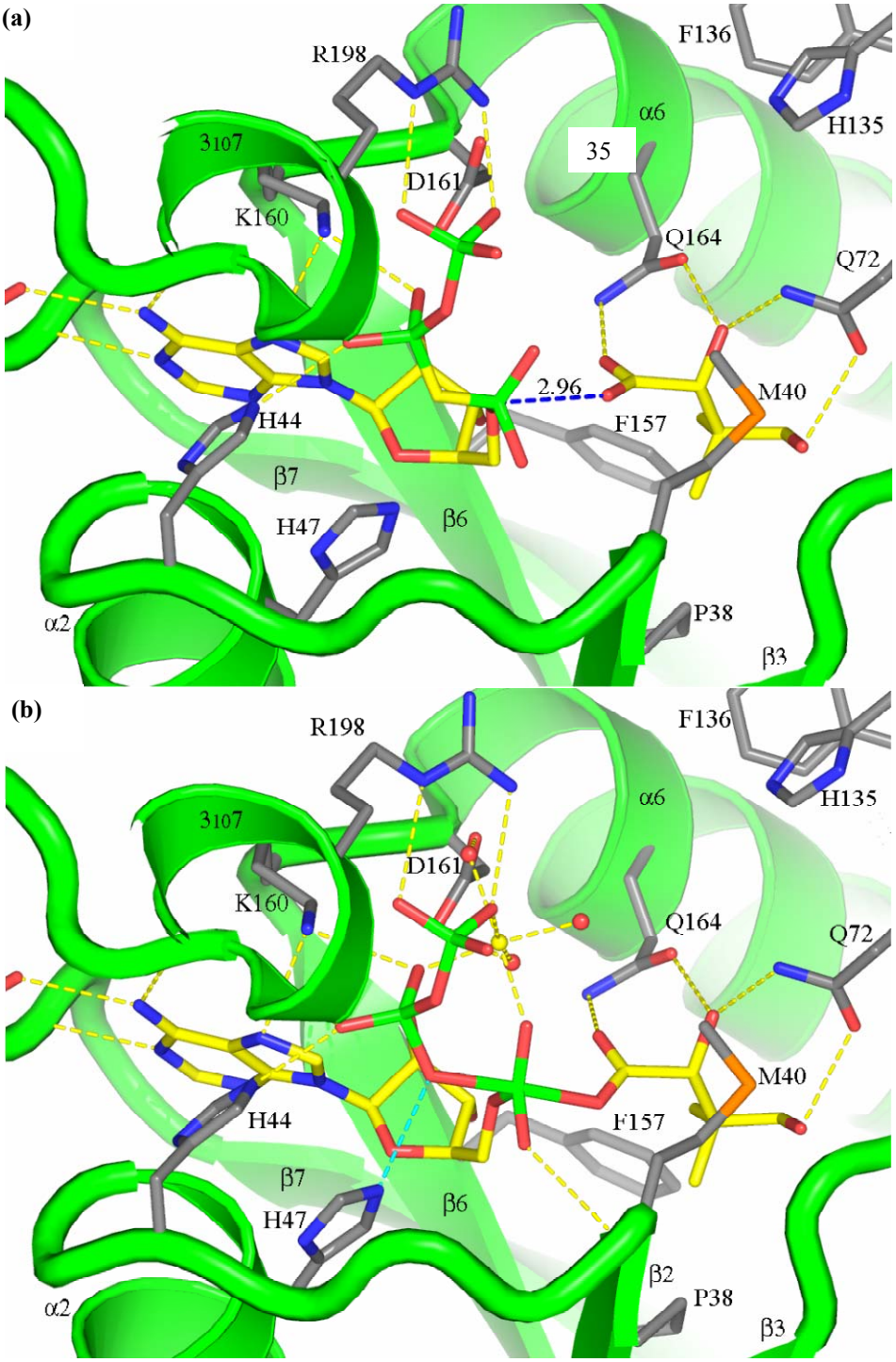



**Figure 4**. **(a)** Binding interactions of AMPCPP and pantoate in the active site of subunit A of the $P2_1$ crystal soaked in a solution containing both AMPCPP and pantoate. Both molecules have a partial occupancy (see text for details), and the magnesium ion is not well defined in this active site. However, binding position and interactions are essentially identical to those of AMPCPP and pantoate with a full occupancy when each is bound in the active site alone. The distance between the α-phosphorus atom and the carboxyl oxygen is ~2.96 Å, shown in blue dashed line. **(b)**

A model of the trigonal bipyramidal intermediate based on crystal structures of the ATP complex, AMPCPP complex and pantoate complex. Formation of the intermediate requires only the movement of the α-phosphorus atom towards the carboxyl oxygen. All favorable interactions with active residues are preserved. His47 is poised to function as a general acid.

### 4.6 Binding Interactions of the Reaction Intermediate, Pantoyl Adenylate

Soaking the $P2_1$ crystals with solutions containing both ATP and pantoate, or crystallization in the presence of both substrates, resulted in a reaction intermediate, pantoyl adenylate, in the active site [10]. Cocrystallization resulted in both active sites in the dimer being fully occupied with the pantoyl adenylate molecule. However, crystals soaked with ATP and pantoate have one active site (subunit B) fully occupied but the other partially occupied, indicated by their difference in B factors. Subunit A of the dimer also has a lower occupancy of AMPCPP when crystals were soaked in solutions containing AMPCPP. This is due to crystal packing, which makes the flexible wall of the active site cavity ($3_{10}$3-loop-α3'-loop region) ordered in subunit A and thus the active site less accessible for bulky molecules ATP and AMPCPP.

Pantoyl adenylate has extensive binding interactions with the active site residues (**Figure 5**). The molecule is almost linear and fits snugly in the bottom of the active site cavity. Both pantoate and adenylate moieties sit in the same positions in the active site as the pantoate molecule of the pantoate complex and the adenosine group of ATP in the ATP complex, and preserve all strong interactions with protein atoms. The tight binding of pantoyl adenylate stabilizes this highly reactive intermediate. One important feature of the pantoyl adenylate complex is that the disordered loop in subunit B becomes ordered, in both cocrystallization and soaking the apo-enzyme crystals with both ATP and pantoate. This suggests that the flexible wall can function as a gate, opening for substrates, especially the bulky ATP molecule, to bind and closing the active site cavity to protect the highly reactive pantoyl adenylate.

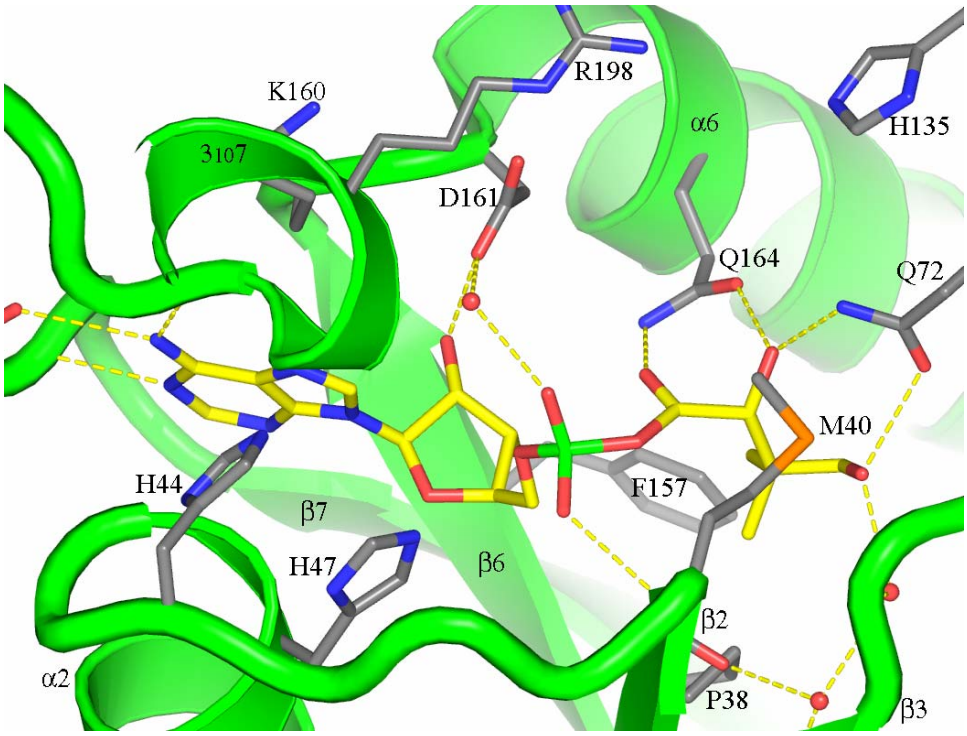



**Figure 5**. Binding interactions of the pantoyl adenylate intermediate. The intermediate binds tightly at the bottom of the active site cavity with many hydrogen bonds and hydrophobic interactions. Hydrogen bonds to the O3* of ribose and some water mediate hydrogen bonds are not shown for clarity.

### 4.7 Binding Interactions of Beta-Alanine

Cocrystallization and soaking crystals with β-alanine [10] failed to yield a complex with this substrate, suggesting that the binding site for β-alanine exists only after pantoyl adenylate is formed in the active site. Based on the structure of the pantoyl adenylate complex, it is likely that the phosphate group of pantoyl adenylate serves as an anchor for the initial binding of β-alanine by offering hydrogen-bonding partners and/or favorable charge-charge interactions. Furthermore, an AMP molecule bound at the active site may also assist the binding of β-alanine from

its phosphate group. Therefore, we soaked crystals of PS in a solution containing both β-alanine and AMP, and we obtained a crystal structure of the PS enzyme in complex with β-alanine and AMP from the $P2_1$ crystals [11].

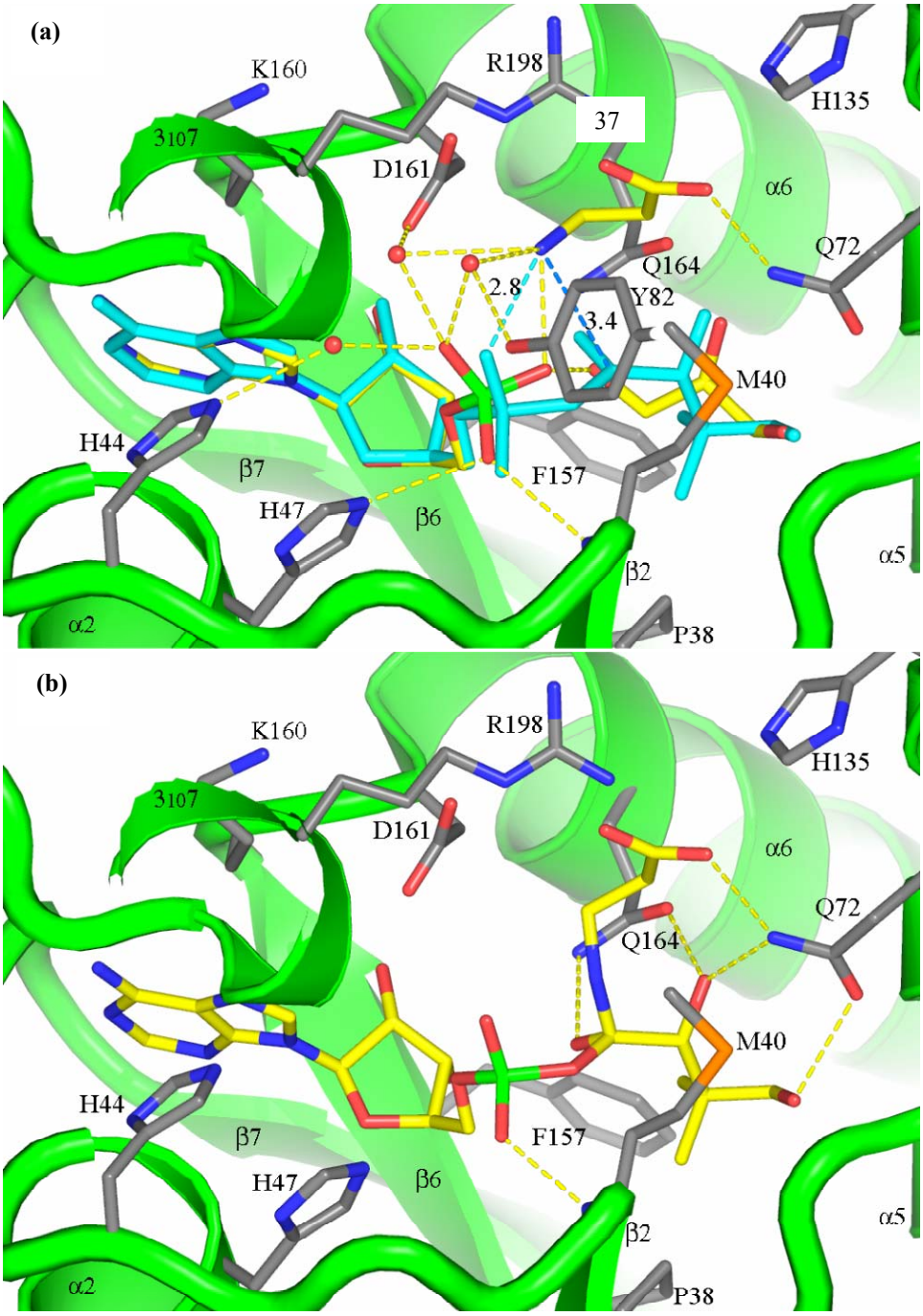



**Figure 6**. **(a)** Structural superposition of the active site of the pantoyl adenylate complex with that of the β-alanine/AMP complex. The active site residues align well with an rmsd of Cα less than 0.15 Å. The pantoyl adenylate molecule is colored in cyan. The protein structure shown in the figure is from the β-alanine/AMP complex. Hydrogen bonds to the phosphate group of AMP and to β-alanine are shown in yellow dashed lines. Some water mediated hydrogen bonds are not shown for clarity. The Tyr82 side chain is from a loop in the front that is clipped away to reveal the bound AMP and β-alanine. Based on the structural alignment, the amino group of β-alanine is ~2.8 Å from the phosphate oxygen of pantoyl adenylate, and ~3.4 Å from the carbonyl carbon, shown in blue dashed lines. **(b)** Model of the tetrahedral intermediate structure based on the structural superposition in (A). The Tyr82 side chain is not shown for clarity. A tetrahedral geometry can be obtained by a torsional rotation of the amino nitrogen of β-alanine to move it to ~1.5 Å above the carbonyl carbon. All favorable interactions with protein atoms are preserved, and thus the intermediate is stabilized by the enzyme.

The β-alanine molecule binds in the active site of subunit A in the upper part of the active site cavity, with its amino group near the phosphate group of AMP (**Figure 6a**). The amino group forms hydrogen bonds to the phosphate group of AMP and to two water molecules. Both water molecules are fixed in position through one hydrogen bond to the phosphate group and a second hydrogen bond to the side chain of Asp161 or Tyr82. The carboxyl group of β-alanine has one hydrogen bond to the Nε2 atom of th[illegible]72 side chain. This carboxyl is also close to the side chains of Arg198 and His135, with distances of ~4.0 [illegible] Å, respectively, indicating favorable charge-charge interactions and π-electron interactions. One side of β-alanine faces side chains of Met40 and Tyr82; the other side faces the large cavity of the active site, which is filled with a few ordered water molecules in the crystal structure. Binding of β-alanine is not as tight as that of other substrates and its binding site in the crystal is not fully occupied. However, the binding interactions limit molecules that can bind well and have a nucleophilic attack on the reaction intermediate. Therefore, structural analogs of β-alanine are all poor substrates [8].

In subunit B of the dimer, where residues 74 to 83 (including Tyr82) are disordered, only AMP but no β-alanine was found in the active site [11]. Also in this case, disorder of this active site cavity wall weakens binding interactions for β-alanine because its binding site is exposed to bulk solvent and thus the charge-charge interactions are dampened. Similarly, in the C2 crystal, which has one molecule per asymmetry unit and the flexible loop disordered, only AMP was found in the active site when crystals were soaked in a solution containing both AMP and β-alanine.

Superposition of the structure of the β-alanine/AMP complex with that of the pantoyl adenylate complex (**Figure 6a**) shows that the amino group of β-alanine was positioned above the carbonyl group of pantoyl adenylate such that the nitrogen atom was ~3.4 Å from the carbonyl carbon. The O1α atom of pantoyl adenylate can form a hydrogen bond with the amino group of β-alanine, serving as an anchor for the initial binding of β-alanine. With changes in only one torsion angle, the amino group of β-alanine can move towards the carbonyl carbon for a nucleophilic attack and attain a nearly perfect geometry for a tetrahedral intermediate (**Figure 6b**).

### 4.8 AMP Binding Interactions

As described above, we obtained AMP complexes from both $P2_1$ and C2 crystals by soaking crystals in a solution containing both AMP and β-alanine. Binding interactions of AMP with the active site residues in these complexes are essentially identical. The adenosine group of AMP binds in the same position as that of the pantoyl adenylate, AMPCPP, and ATP: it fits snugly in its binding pocket at the bottom of the active site cavity through hydrophobic and hydrogen bonding interactions. The adenine group is flanked by Gly46 on helix α2 and Lys160 on the loop after β6. Its N1 and N6 atoms have hydrogen bonds to main chain atoms. The hydroxyl groups of ribose form hydrogen bonds with the Asp161 side chain and a few main chain atoms at the bottom of the active site cavity. The phosphate group, however, has torsional flexibility, and it rotates slightly relative to the α-phosphate of ATP, allowing one of its oxygen atoms to form hydrogen bonds to the side chain of His47 and to the amide nitrogen of Met40 simultaneously (**Figure 6a**).

## 5. OVERALL ENZYME-CATALYZED REACTION MECHANISM

Putting all structural data together allows us to have a detailed view of the overall enzyme-catalyzed reaction mechanism. When the active site is empty, the disordered loop opens the active site and thus allows the ATP molecule to come in and bind. Because ATP is relatively bulky and is held rigidly in the active site by many binding interactions, it is likely that ATP binds first. The smaller sized pantoate then comes in, and initiates a nucleophilic attack on the α-phosphate while it binds in its binding pocket [8, 24]. When both ATP and pantoate are present in the same active site, the nucleophilic reaction must occur (**Figure 4a**). Formation of the transient trigonal bypyramidal intermediate involves only the movement of the phosphate atom towards the carboxyl oxygen of pantoate to be coplanar with the three equatorial oxygen atoms (**Figure 4b**). Although transient in nature, this intermediate is stabilized by the enzyme because all the favorable interactions are preserved, and the unfavorable steric hindrance and charge repulsions between ATP and pantoate are eliminated. The trigonal bipyramidal intermediate then dissociates, and simultaneously the pyrophosphate group leaves the active site. The magnesium ion and the positively charged side chains around the β- and γ-phosphate groups draw the negative charges towards the leaving pyrophosphate. His47 is also likely to facilitate the reaction by donating a proton to the leaving pyrophosphate. This leaves a reaction intermediate, pantoyl adenylate in the active site and completes the first half of the enzyme catalyzed reaction. Pantoyl adenylate is stabilized by many strong binding interactions, as we have seen earlier. It is also protected by the closing of the active site when the flexible loop becomes ordered. Pantoyl adenylate is stable only when tightly bound in the enzyme active site. In solution, it decomposes rapidly to give

pantoyl lactone and AMP [9]. The dissociation of the trigonal bipyramidal intermediate necessitates the simultaneous leaving of the pyrophosphate, because there would be unfavorable charge and steric repulsions once pyrophosphate is formed. This reaction, however, is reversible, as was demonstrated by positional isotope exchange experiments that $^{18}O$ labels scrambled when incubating [$\beta\gamma$-$^{18}O_6$]-ATP and pantoate with the MTB PS enzyme [24].

The presence of pantoyl adenylate intermediate and the ordering of the flexible wall create the binding site for β-alanine, the last substrate. The β-alanine molecule is relatively small, capable of passing through the small opening at the top of the active site cavity. The phosphate group of pantoyl adenylate can have a hydrogen bond to the amino group of β-alanine, thus serving as an initial anchor for binding β-alanine. The carboxylate group of β-alanine has favorable interactions with the side chains of Arg198 and His135 and a hydrogen bond to Gln72. There is a shift of the Arg198 side chain of ~1.8 Å towards the carboxylate group of β-alanine relative to that in the pantoyl adenylate complex. This binding position for β-alanine puts its amino group at a good position for a nucleophilic attack on the carbonyl carbon of the pantoyl group. The distance of the amino nitrogen is only ~3.4 Å from the carboxyl carbon of pantoyl adenylate. A model of the tetrahedral intermediate can be readily prepared from the crystal structures of the pantoyl adenylate complex and the β-alanine/AMP complex (**Figure 6b**). Formation of the tetrahedral intermediate requires only a torsional rotation of the amino nitrogen atom of β-alanine toward the carbonyl carbon of pantoyl adenylate. Similar to that of the trigonal bypyramidal intermediate, formation of the tetrahedral intermediate preserves all favorable binding interactions of active site residues with pantoyl adenylate and β-alanine. Therefore, the intermediate is stabilized by the enzyme. Dissociation of the tetrahedral intermediate then forms pantothenate and AMP. The planarity of the peptide bond of pantothenate causes significant rearrangement of the molecule, and several hydrogen bonds to active site residues are broken. In addition, the newly formed pantothenate has steric clashes and charge repulsions with AMP and some active site residues. Therefore, pantothenate must leave the active site once it is formed. The AMP molecule then diffuses away to regenerate the enzyme active site for next cycle of catalysis. Pantothenate has very low binding affinity to the enzyme. We have tried soaking and growing crystals in the presence of pantothenate, but we failed to obtain a complex with this compound. This is consistent with the finding that pantothenate is a poor inhibitor of the enzyme [8]. The AMP molecule, on the other hand, has good binding interactions in the active site. However, it typically has a low concentration inside cells, and thus can easily diffuse out of the active site. Moreover, ATP has many favorable interactions in the active site of the enzyme, and it can readily displace AMP.

## 6. REFERENCES


[1] S. Jackowski. Escherichia coli and Salmonella typhimurium: Cellular and molecular biology, American Society for Microbiology, Washington, DC 1996, pp. 687-694.

[2] W.K. Maas, Proc Int Congr Biochem 11 (1960) 161-168.

[3] V.K. Sambandamurthy, X. Wang, B. Chen, R.G. Russell, S. Derrick, F.M. Collins, S.L. Morris and W.R. Jacobs, Jr., Nat Med 8 (2002) 1171-4.

[4] W.K. Merkel and B.P. Nichols, FEMS Microbiol Lett 143 (1996) 247-52.

[5] F. von Delft, A. Lewendon, V. Dhanaraj, T.L. Blundell, C. Abell and A.G. Smith, Structure (Camb) 9 (2001) 439-50.

[6] U. Genschel, C.A. Powell, C. Abell and A.G. Smith, Biochem J 341 ( Pt 3) (1999) 669-78.

[7] A. Perez-Espinosa, T. Roldan-Arjona and M. Ruiz-Rubio, Mol Genet Genomics 265 (2001) 922-9.

[8] R. Zheng and J.S. Blanchard, Biochemistry 40 (2001) 12904-12.

[9] T. Wieland, W. Loewe, A. Kreiling and G. Pfleiderer, Biochem Z 339 (1963) 1-7.

[10] S. Wang and D. Eisenberg, Protein Sci 12 (2003) 1097-108.

[11] S. Wang and D. Eisenberg, Biochemistry (2005) In press.

[12] Z. Otwinowski and W. Minor, Methods Enzymol 276 (1996) 307-326.

[13] J. Navaza, Acta Crystallogr A 50 (1994) 157-163.

[14] A.T. Brunger, P.D. Adams, G.M. Clore, W.L. DeLano, P. Gros, R.W. Grosse-Kunstleve, J.S. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson and G.L. Warren, Acta Crystallogr D Biol Crystallogr 54 ( Pt 5) (1998) 905-21.

[15] G.N. Murshudov, A.A. Vagin, A. Lebedev, K.S. Wilson and E.J. Dodson, Acta Crystallogr D Biol Crystallogr 55 ( Pt 1) (1999) 247-55.

[16] T.A. Jones, J.Y. Zou, S.W. Cowan and Kjeldgaard, Acta Crystallogr A 47 ( Pt 2) (1991) 110-9.

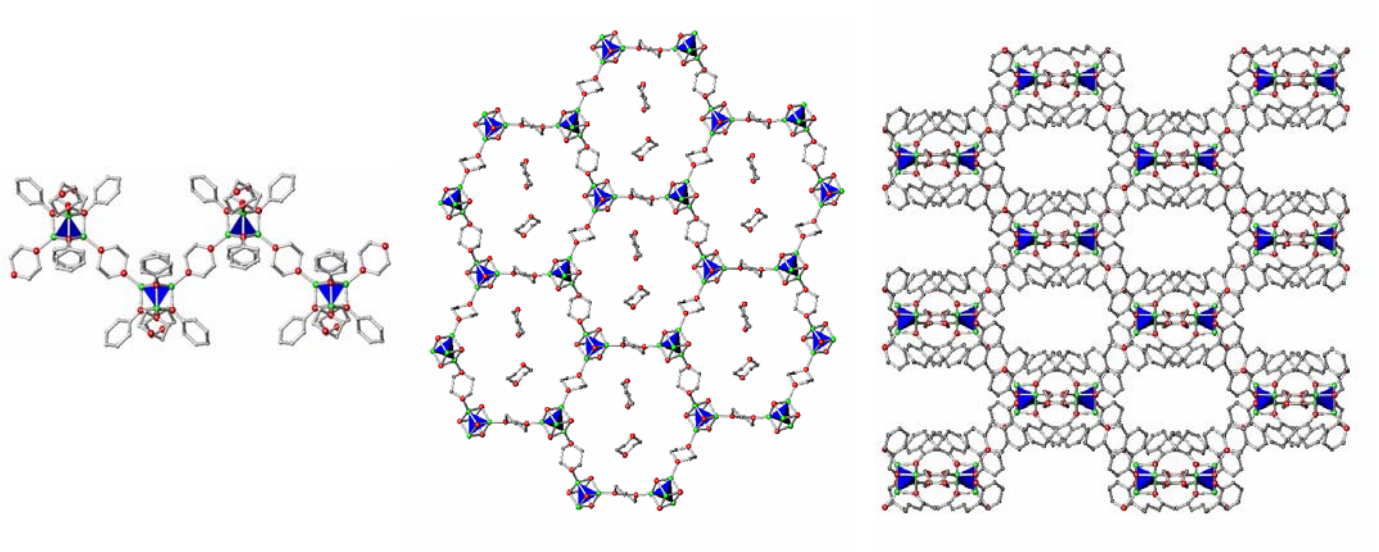
[17] P.J. Kraulis, Journal of Applied Crystallography 24 (1991) 946-950.

[18] E.A. Merritt and M.E. Murphy, Acta Crystallogr D Biol Crystallogr 50 (1994) 869-73.

[19] D.G. Barker and G. Winter, FEBS Lett 145 (1982) 191-3.
[20] W.L. DeLano, The PyMOL molecular graphics system, DeLano Scientific, San Carlos, CA, 2002.
[21] P. Bork, L. Holm, E.V. Koonin and C. Sander, Proteins 22 (1995) 259-66.
[22] R. Zheng, T.K. Dam, C.F. Brewer and J.S. Blanchard, Biochemistry 43 (2004) 7171-8.
[23] C. Hountondji, P. Dessen and S. Blanquet, Bio 40 e 68 (1986) 1071-8.
[24] L. Williams, R. Zheng, J.S. Blanchard and F.M. Raushel, Biochemistry 42 (2003) 5108-13.

*Proceedings from the Symposium*

# Margaret C. Etter Early Career Award



## ETTER TRANSACTIONS OF THE AMERICAN CRYSTALLOGRAPHIC ASSOCIATION

Edited by Frances Jurnak

VOLUME 1, 2005

*Proceedings from the Symposium*

# Margaret C. Etter Early Career Award

ACA ETTER TRANSACTIONS
VOLUME 1

Walt Disney World, Orlando, Florida
May 28 – June 2, 2005

**Editor**

Frances Jurnak
University of California, Irvine
Tel: 949-824-6623
Fax: 949-824-8540
jurnak@uci.edu

# BIS-CHELATION AND ANION EFFECTS INVOLVING A MOLECULE CONSTRUCTED IN THE ORGANIC SOLID STATE USING MOLECULAR TEMPLATES

Tamara D. Hamilton, Giannis S. Papaefstathiou and Leonard R. MacGillivray
*University of Iowa Department of Chemistry, Iowa City, IA 52242*

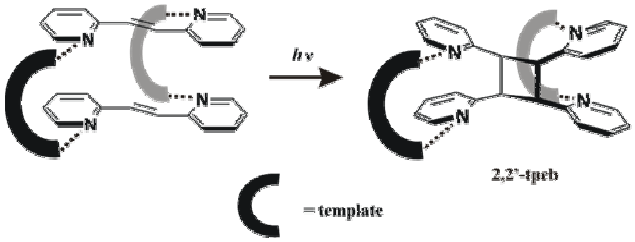
## 1. ABSTRACT

A molecule derived from a template-directed solid-state synthesis, *rctt*-tetrakis(2-pyridyl)cyclobutane (2,2'-tpcb), serves as a bis-chelating ligand upon reaction with $Cu(BF_4)_2{\cdot}H_2O$ and $NEt_4PF_6$ to produce the dinuclear complex $[Cu_2(OH)_2(H_2O)_4(\mu\text{-}2,2'\text{-tpcb})][PF_6]_2$ (**1**).

## 2. INTRODUCTION

Transition-metal-ion complexes involving seven-membered chelation rings have emerged as important for applications in catalysis and biology. Specifically, N-$C_4$-N bridged metallocycles [1] have found applications as catalysts [2] and models of biological compounds [3], as well as antibacterial [4] and antitumor [5] agents. In terms of catalysis, the application of chelating N-donor ligands - as compared to analogous polyphosphines - has begun to attract attention, especially for the activation of C-H bonds [6]. Thus, ligands that provide more than one pyridyl group in close proximity are being studied as preorganized metal chelation units [6b]. In addition, there is increasing interest in the development of novel non-metallocene catalysts for olefin polymerization for which complexes involving N-$C_4$-N bridged chelate rings have been effective [2c]. Therefore, studies that involve N-$C_4$-N ligands with a potential to offer seven-membered chelation are of much current interest.

The ability to position molecules in an orientation appropriate for reaction in the organic solid state using molecular templates that operate *via* hydrogen bonds has led to the quantitative and gram-scale construction of molecules difficult to obtain from solution [7,8]. In particular, we have shown that templates based on resorcinol can orient olefins [*e.g. trans*-1,2-bis(2-pyridyl)ethylene) (2,2'-bpe)] in the solid state in positions suitable for intermolecular [2+2] photodimerizations. The suitability of the products as ligands for coordination chemistry is evidenced by the presence of pyridine rings that emanate from central cyclobutane rings.




**Scheme 1.** Template-directed solid-state synthesis of 2,2'-tpcb from 2,2'-bpe.

As part of an ongoing study to elucidate anion effects involving transition-metal-ion complexes of *rctt*-tetrakis(2-pyridyl)cyclobutane (2,2'-tpcb), we wish to report here the synthesis and structure determination of the dinuclear Cu(II) complex $[Cu_2(OH)_2(H_2O)_4(\mu\text{-}2,2'\text{-tpcb})][PF_6]_2$ (**1**). In line with our previous study [7], the tetrapyridine acts as a bis-chelating bridge with two Cu(II) centers, giving rise to two seven-membered metallocycles. In contrast to our previous work, the $PF_6^-$ counter ions are non-coordinating and, thus, provide an opportunity for each Cu(II) ion to interact with additional donor units. In the case of **1**, each Cu(II) center is coordinated by two water molecules and a single hydroxide anion, each of which participates in hydrogen bonds with the $PF_6^-$ anion.

## 3. METHODS

Complex **1** was prepared by combining methanolic solutions of 2,2'-tpcb (0.025 g, 0.068 mmol) [8a] and $Cu(BF_4)_2{\cdot}H_2O$ (0.032 g, 0.137 mmol) (1:2 ratio). To the resulting clear blue solution was added a methanolic solution of $NEt_4PF_6$ (0.037 g, 0.137 mmol) (25 mL methanol total). A 5 mL aliquot of the latter solution was then

layered with 10 mL benzene in a test tube and allowed to sit for a period of approximately one week. The solution produced blue single crystals (yield: 0.015 g, 29 %) suitable for X-ray analysis [9].

## 4. RESULTS AND DISCUSSION

Compound **1** crystallizes in the space group $P2_1/n$. An ORTEP perspective of the asymmetric unit of **1** is shown in **Figure 1a**. The asymmetric unit consists of one Cu(II) ion, one-half molecule of 2,2'-tpcb, two water molecules, one hydroxide ion, and one $PF_6^-$. As shown in **Figure 1b**, the dinuclear complex, which sits around a crystallographic center of inversion, contains one molecule of 2,2'-tpcb that bridges two Cu(II) centers [d(Cu1···Cu1a) 7.47 Å]. The Cu(II) ions of the complex adopt an *anti* conformation and exhibit a bite angle of 88.8° [10]. Similar to $[Cu_2(NO_3)_4(\mu\text{-}2,2'\text{-tpcb})]$ and $[Cu_2(\mu_2\text{-}SO_4)_2(\mu\text{ -}2,2'\text{-tpcb})(H_2O)_2]_\infty$, each seven-membered chelation ring adopts a boat conformation. The boat conformation also compares favorably to the chelation complexes of *bis*(2-pyridyl)ethane with Pt(II) [11,12] and 1,2-bis(6-methylpyridin-2-yl)ethane with Pd(II) [13]. The coordination geometry around each Cu(II) center of **1** is square pyramidal. Two pyridyl units of 2,2'-tpcb and two water molecules (O1 and O2) form the base of the pyramid while a hydroxide ion (O3) occupies the apical position. The metal-ligand distances around the base of the pyramid range from 1.96-2.02 Å, while the Jahn-Teller distorted apical distance is 2.21 Å (Table 1).

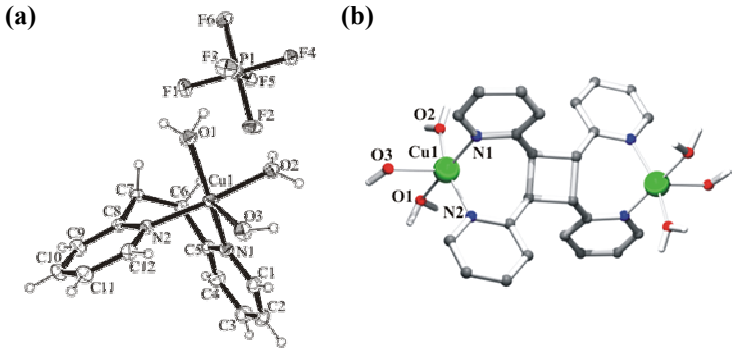



**Figure 1**. **(a)** ORTEP representation of the asymmetric unit of **1**. Displacement ellipsoids are drawn at the 30% probability level, with hydrogen atoms being assigned an arbitrary radius; **(b)** The dinuclear complex **1** (color scheme: green = copper, red = oxygen, blue = nitrogen, grey = carbon). Hydrogen atoms of 2,2'-tpcb have been omitted.

Views of the extended structure of **1** are shown in **Figures 2** and **3**. The complexes have assembled to form layers, parallel to the *ac*-plane, which exhibit an *abab* stacking pattern. The layers are held together by face-to-face π-π forces involving the ligated pyridyl groups (**Figure 2**). The counter $PF_6^-$ ions form layers sandwiched between the layers of Cu(II) complexes. The anions participate in extensive O-H···F hydrogen bonds that involve the two water molecules (O1, O2) and hydroxide ion (O3). Specifically, each $PF_6^-$ anion participates in a total of six O-H···F hydrogen bonds (**Figure 3a**) [d(O1···F3) 2.750(5) Å; d(O1···F4) 2.626(5) Å; d(O2···F2) 2.587(5) Å; d(O2···F6) 2.600(5) Å; d(O3···F1) 2.869(5) Å; d(O3···F5) 2.665(5) Å] (**Figure 3b**) with the ligands.

**TABLE 1**

Selected Bond Lengths (Å) and Angles (°) for **1**

| | | | |
|---|---|---|---|
| Cu1-O1 | 1.987(4) | O1-Cu1-O2 | 84.7(2) |
| Cu1-O2 | 1.963(4) | O1-Cu1-O3 | 99.7(2) |
| Cu1-O3 | 2.214(4) | O1-Cu1-N2 | 97.5(1) |
| Cu1-N1 | 2.020(4) | O2-Cu1-O3 | 83.7(2) |
| Cu1-N2 | 1.986(4) | O2-Cu1-N1 | 92.8(2) |
| | | O3-Cu1-N1 | 101.7(2) |
| | | O3-Cu1-N2 | 100.5(1) |
| | | N1-Cu1-N2 | 88.8(1) |

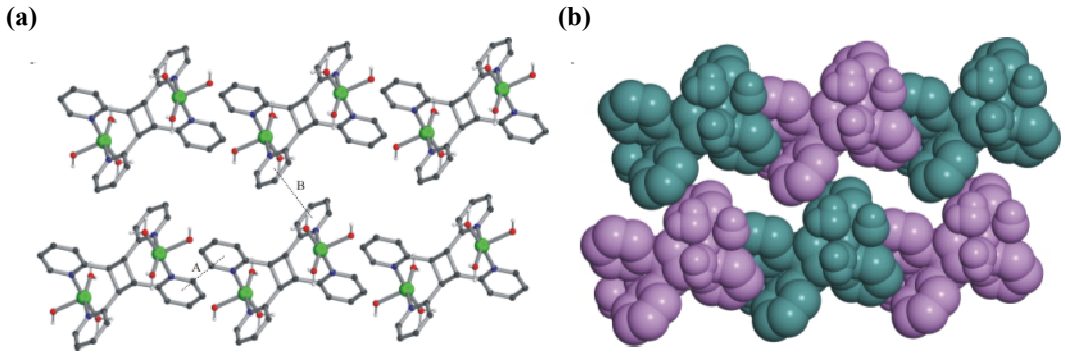



**Figure 2.** Extended structure of **1** showing the π-stacking of pyridine rings viewed in the *ac* plane: **(a)** ball-and-stick representation ($d_A$ = 3.697(5); $d_B$ = 3.874(5) Å) (color scheme: green = copper, red = oxygen, blue = nitrogen, grey = carbon) and **(b)** space-filling representation (hydrogen atoms of 2,2'-tpcb omitted for clarity).

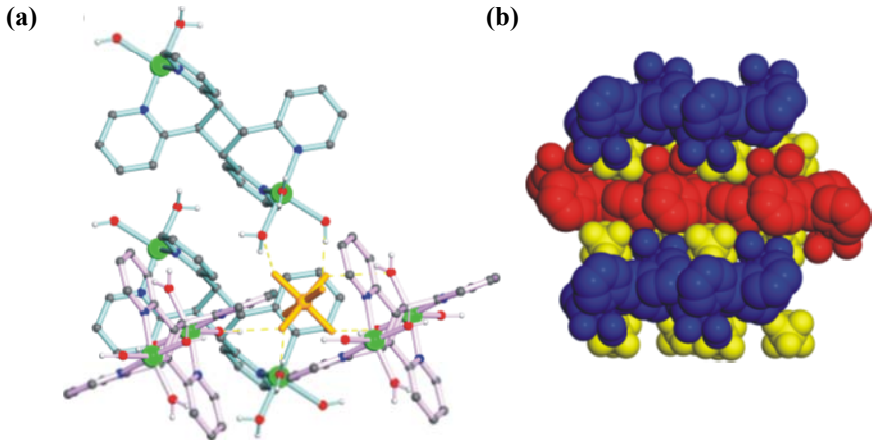



**Figure 3.** **(a)** Hydrogen bonding array involving $PF_6^-$ anion (gold) in **1**. Each F-atom is involved in one hydrogen bond (yellow) to a coordinated water or hydroxide molecule of a neighboring assembly (color scheme: green = copper, red = oxygen, blue = nitrogen, grey = carbon, white = hydrogen) (hydrogen atoms of 2,2'-tpcb omitted for clarity) and **(b)** space-filling representation of **1**, showing *abab* packing of complex layers alternating with $PF_6^-$ anions. Complexes are shown in red or blue, while $PF_6^-$ anions are shown in yellow.

## 5. CONCLUSION

In conclusion, template-directed solid-state organic synthesis has led to the construction of a tetrapyridyl ligand that forms the dinuclear coordination complex **1** with Cu(II) ions. The complex incorporates two seven-membered chelation rings [7]. 2,2'-tpcb is the first ligand to sustain seven-membered chelation in a series of coordination complexes. The non-coordinating behavior of the $PF_6^-$ counter-ion provides open sites on each metal, which are occupied by water molecules and hydroxide ions.

## 6. REFERENCES

[1] a) McCullough, F. Jr.; Bailar, J. C. Jr. *J. Am. Chem. Soc.* **1956**, *78*, 714-716;
b) Habu, T.; Bailar, J. C. Jr. *J. Am. Chem. Soc.* **1966**, *88*, 1128-1130.

[2] a) Gibson, V. C.; Kimberley, B. S.; White, A. J. P.; Williams, D. J.; Howard, P. *Chem. Commun.* **1998**, 313-314;
b) Kretzschmar, E. A.; Kipke, J.; Sundermeyer, J. *Chem. Commun.* **1999**, 2381-2382;
c) Siemeling, U.; Kolling, L.; Stammler, A.; Stammler, H. G.; Kaminski, E.; Fink, G., *Chem. Commun.* **2000**, 1177-1178;
d) Bacchi, A.; Pelizzi, G.; Minardi, G.; Pistuddi, A. M.; Solinas, C.; Chelucci, G. *Trans. Met. Chem.* **2002**, *27*, 274-278;
e) Minardi, G.; Mura, E.; Pistuddi, A. M.; Solinas, C.; Bacchi, A.; Pelizzi, C.; Pelizzi, G.; Chelucci, G *Trans. Met. Chem.* **1999**, *24*, 481-485;

f) Zhu, H. B.; Dai, Z.-Y.; Huang, W.; Cui, K.; Gou, S.-H.; Zhu, C.-J. *Polyhedron* **2004**, *23*, 1131-1137;
g) Bandoli, G.; Gerber, T. I. A.; Jacobs, R.; du Preez, J. G. H. *Inorg. Chem.*, **1994**, *33*, 178-179.

[3] a) Martin, J. W. L.; Timmons, J. H.; Martell, A. E.; Rudolf, P.; Clearfield, A. *Inorg. Chem.* **1981**, *20*, 814-821;
b) Choo P. L.; Mulichak, A. M.; Jones Jr., R. W.; Bacon, J. W.; Pett, V. B.; Zacharias, D. E. *Inorg. Chim. Acta* **1990**, *171*, 183-192;
c) Ball, R. G.; Trotter, J. *Can. J. Chem.* **1979**, *57*, 1368-1371;
d) McFarland, C. A.; Gross, S. A.; Winfield, J. S.; Pett, V. B. *Inorg. Chim. Acta*, **1994**, *221*, 35-41;
e) Balamurugan, R.; Palaniandavar, M.; Gopalan, R. S. *Inorg. Chem.* **2001**, *40*, 2246-2255.

[4] Nandi, A. K.; Chaudhuri, S.; Mazumdar, S. K. Ghosh, S. *Inorg. Chim. Acta* **1984** *92*, 235-240.

[5] a) Khokar, A. R.; Shamsuddin, S.; Xu, Q. *Inorg. Chim. Acta* **1994**, *219*, 193-197;
b) Shamsuddin, S.; van Haal, J. W.; Stark, J. L.; Whitmore, K. H.; Khokhar, A. R. *Inorg. Chem.* **1997**, *36*, 5969-5971;
c) Shamsuddin, S.; Santillan, C. C.; Stark, J. L.; Whitmire, K. H.; Siddik, Z. H.; Khokhar, A. R. *J. Inorg. Biochem.* **1998**, *71*, 29-35;
d) Hoeschele, J. D.; Showalter, H. D. H.; Kraker, A. J.; Elliott, W. L.; Roberts, B. J.; Kampf, J. W. *J. Med. Chem.* **1994**, *37*, 2630-2636.

[6] a) Vedernikov, A. N.; Huffman, J. C.; Caulton, K. G. *Inorg. Chem.* **2002**, *41*, 6244-6248;
b) Vedernikov, A. N.; Huffman, J. C.; Caulton, K. G. *Inorg. Chem.* **2002**, *41*, 6867-6874; Vedernikov, A. N.; Pink, M.; Caulton, K. G. *Inorg. Chem.* **2004**, *43*, 3642-3646.

[7] Hamilton, T. D.; Papaefstathiou, G. S.; MacGillivray, L. R. *Cryst. Eng. Comm.* **2002**, *4*, 223-226.

[8] a) Papaefstathiou, G. S.; Kipp, A. J.; MacGillivray, L. R. *Chem. Commun.* **2001**, 2462-2463;
b) MacGillivray, L. R.; Reid, J. L.; Ripmeester, J. A. *J. Am. Chem. Soc.* **2000**, *122*, 7817-7818;
c) Hamilton, T. D.; Papaefstathiou, G. S.; MacGillivray, L. R. *J. Am. Chem. Soc.* **2002**, *124*, 11606-11607.;
d) Varshney, D. B.; Papaefstathiou, G. S.; MacGillivray, L. R. *Chem. Commun.* **2002**, 1964-1965;
e) Gao, X.; Friščić, T.; MacGillivray, L. R. *Angew. Chem. Int. Ed.* **2004**, *43*, 232-236.

[9] Crystal data for **1**: monoclinic, $P2_1/n$, Z = 4, a = 9.716(5), b = 16.586(5), c = 9.939(5) Å, $\beta$ = 98.270(5) º, V = 1585.0(12) Å$^3$, 2738 unique reflections, $R$ = 0.0492 [for 2122 reflections with $I \geq 2\sigma(I)$], $R$ = 0.0742 and w$R^2$ = 0.1408 (all data). Crystallographic Data has been deposited with the Cambridge Crystallographic Data Centre and has been allocated with deposition number CCDC 291578.

[10] The bite angle is the angle between chelating ligand donor atoms (L-M-L) in a metal complex.

[11] McFarlane, A.; Lusty, J. R.; Fiol, J. J.; Terron, A.; Molins, E.; Miravitlles, C.; Moreno, V. *Z. Naturforsch., Teil B*, **1994**, *49*, 844-848.

[12] a) Marcelis, A. T. M.; Korte, H.-J.; Krebs, B.; Reedijk, J. *Inorg. Chem.*, **1982**, *21*, 4059-4063;
b) Kleibohmer, W.; Krebs, B.; Mercelis, A. T. M.; Reedijk, J.; Van der Veer, J. L. *Inorg. Chim. Acta* **1983**, *75*, 45-50.

[13] Baker, A. T.; Crass, J. K.; Kok, G. B.; Orbell, J. D.; Yuriev, E. *Inorg. Chim. Acta* **1993**, *214*, 169-176.

# USE OF ALKALI METAL AGGREGATES IN CONTROLLING NETWORK ASSEMBLY

Dugald J. MacDougall, J. Jacob Morris, Bruce C. Noll and Kenneth W. Henderson*
*Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, Indiana 46556, USA.*

## 1. ABSTRACT

This paper summarizes the first systematic studies into the use of *s*-block aggregates to control the rational assembly of network polymers. Preliminary work using lithium and sodium aryloxide aggregates has established the efficacy of this methodology, and in particular that the control of network architectures can be achieved on a regular basis. These studies also demonstrate that air-stable, robust, open-framework materials may be prepared through use of these systems. Overall, the paper outlines an area that has considerable potential for the preparation of new materials, that advance knowledge in the understanding of the self-assembling process.

## 2. INTRODUCTION

The synthesis and characterization of periodic network assemblies is an area of increasing international academic and industrial interest [1-4]. The attention paid to this topic can be attributed, in part, to the fundamental scientific challenges present in developing rational routes to structurally well-defined solids and also as a consequence of the potential utility of these materials in applications as diverse as catalysis [5], chemical separation [6], optics [7] and electronics [8]. In this regard, two main strategies have been adopted in the formation of frameworks from molecular precursors [9], firstly through the use of hydrogen-bonded organic solids [10] and secondly using metal-organic coordination polymers containing transition elements and ligand spacer molecules [11]. These strategies have also been combined where metal-containing complexes are incorporated within hydrogen-bonded networks [12]. In addition, a notable area of success in controlling the formation of covalently-linked solids has been the use of metal carboxylate clusters as directional components of metal-organic frameworks (MOFs) [13]. However, a prominent omission from the current list of strategies directed towards rational network synthesis is the use of early main group metals to dictate supramolecular structure. This deficit may be explained by the considerable potential problems associated with these species, such as the presence of complex dynamic solution equilibria between aggregated complexes; the spherical non-polarizable nature of the cations; the limited preference for fixed metal geometries, and the lability of the metal-ligand bonding. All of these issues may lead to poor predictability and reproducibility in network synthesis [14-16]. In this context it is unsurprising that researchers have opted for alternative structural scaffolds, and to a large extent have consciously avoided the use of *s*-block metal cations. In this paper we directly address these established perceptions and contest that the early main group elements can be successfully employed to control network assembly by the informed selection of metal, anionic partner, linker molecule and synthetic conditions. More specifically, we outline a novel approach to rational network synthesis utilizing pre-assembled *s*-block molecular aggregates as secondary building units (SBUs) to control supramolecular structure. Although the use of SBUs is quickly emerging as a successful tactic in rationalizing and controlling the formation of various types of extended framework architectures, the use of *s*-block SBUs has not been investigated [17]. In this paper we detail our preliminary investigations in this area, which firmly establish the efficacy of this methodology. We will outline an unprecedented systematic series of one-, two- and three-dimensional polymers constructed from *s*-block metal aggregates. A subset of these networks displays intriguing host-guest chemistry, including the preparation of air-stable, robust, open-framework materials. We have also developed a model for network prediction, which will be an essential element in our future studies this area.

### 2.1 Initial Observations and Proof-of-Concept Preliminary Studies

During the course of our investigations into the structure and function of $\alpha$, $\alpha$'-stabilized carbanions we discovered that the lithium salts of many of these species form unusual polymeric materials [18]. We structurally characterized lithiated $\alpha$-cyanophosphonates of the type $[(RO)_2P(O)CHCNLi.THF]$ (R = Et or $Pr^i$) and found that they form 2D sheet structures, as displayed in **Figure 1**.

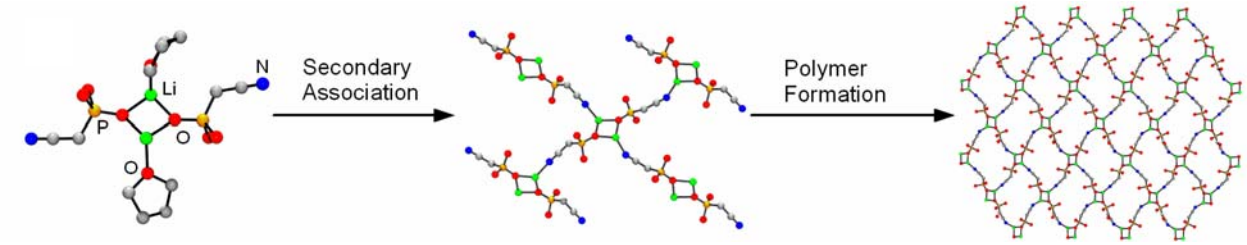



**Figure 1.** Assembly process of lithiated α-cyanophosphonates by linking $Li_2O_2$ dimers.

Intriguingly from the molecular perspective, the sheets can be described as being composed of 'dimeric' $Li_2O_2$ units that are interconnected to four identical 'dimers' via ligation of the nitrile groups. The $Li_2O_2$ motif, where the metals bridge between a pair of phosphoryl units, is entirely consistent with established aggregation patterns for simple lithiated phosphonates [19]. Then, the pendant nitriles act as linear-linkers and bond in an 'interdimer' fashion to another $Li_2O_2$ unit, with the adjacent rings rotated by approximately 90° with respect to one another. Overall, four dimers interconnect through the nitrile bridges to give large $(LiOPCCN)_4$ twenty-four membered rings with small four membered $Li_2O_2$ rings at the corners. This pattern is topologically equivalent to a (4,4) net and can be described as a decorated network where one square-planar vertex has been replaced by a molecular aggregate [20, 21]. The networks obtained are reminiscent of classical structures utilizing square-planar transition metal nodes, such as $Ni^{2+}$, $Zn^{2+}$, $Cu^{2+}$, $Cd^{2+}$, *etc.*, with neutral linear-linker ligands such as 4,4'-bipyridine. In our case the $Li_2O_2$ dimeric aggregates act as square-planar nodes, where the nodal point is located at the center of the ring rather than being an individual atom.

We theorized that it should be possible to form related architectures in a rational and controlled manner using other types of lithiated aggregates. We chose to examine (organo)sulfonylacetonitriles, $RSO_2CH_2CN$, as carbanion sources since simple lithiated sulfonyls are known to form $(SO_2Li)_2$ eight membered ring dimers [22]. Therefore, these complexes should assemble in a similar manner to the cyanophosphonates described above but with larger eight membered rings interconnecting the macromolecular twenty-four membered rings. In this vein we targeted and subsequently successfully prepared and structurally characterized the complex $[MeSO_2CHCNLi.THF]$, [23] and we were pleased to confirm that this complex forms a network with exactly the connectivity that was predicted in advance [24, 25]. **Figure 2** illustrates the formation of this two dimensional sheet that may be described as a basket weave network. Overall, the $(SO_2Li)_2$ eight-membered ring acts as an offset square planar node to give the desired topology.

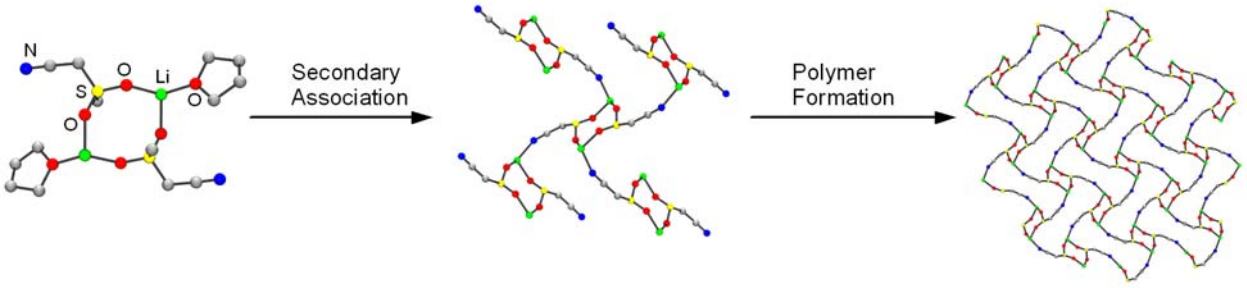



**Figure 2.** Assembly process of lithiated (organo)sulfonylacetonitriles through association of $(SO_2Li)_2$ dimers.

## 3. METHODS

The coordination chemistry involving the early main group elements is an active area of academic study, with the molecular aggregation behavior of numerous classes of metal-ligand combinations now well established [14-16]. With this understanding of molecular aggregation in hand, we aimed to select a series of readily prepared and well-behaved main group aggregates that could then be utilized as SBUs. A highly attractive feature of *s*-block SBUs is the diversity of metal aggregate geometries and compositions available as synthons. Our goal was to utilize the

geometric and chemical information stored in a variety of metal-containing ring and cage compounds to rationally design, and subsequently build, novel two- and three-dimensional network architectures. Specifically, since the location of the metal centers and the anions within these aggregates are known in advance, we could use this information to predict the directions for polymer extension. Suitable aggregates were selected that contain metals with one or more free coordination sites or labile ligands for Lewis base ligation and subsequent polymer growth. In essence the entire aggregate is used as a fixed nodal point (as opposed to the nodes being single spherical metal cations) with their geometries determined by the nature of the rings or cages. This strategy includes utilizing strong 'primary' bonding interactions to assemble molecular aggregates of predictable composition, shape and size, which are then linked via weaker 'secondary' Lewis base-type bonding. The interaggregate association is achieved by the use of neutral polydentate Lewis bases containing at least two donor sites, as opposed to the use of internal donor groups attached to the anionic ligand backbones as described in **Figures 1** and **2**. The use of SBUs in network synthesis is also an appealing approach since they dramatically reduce the number of possible network topologies arising for a given node/linker combination and they may also aid in the preparation of stable porous materials [26].

It is pertinent to recognize that polymer formation is in fact very common for early main group compounds, in particular for *s*-block complexes [27, 28]. Indeed, most preparative chemists have from time to time encountered an insoluble lithiated intermediate during a synthesis. In many instances such solubility problems can be overcome by the use of polar solvents such as THF, DME and TMEDA, which cleave metal-ligand interactions and generally reduce the aggregation state of the metallated species. In turn, the vast majority of work carried out on the structure and bonding of early main group metal complexes has been performed at the molecular level, and the structural elucidation of polymeric materials has mainly been by inference rather than by direct analysis [29]. Although, various types of coordination polymers containing *s*-block metals have previously been characterized, they have usually been prepared either inadvertently or in order to study localized metrical information, rather than deliberately to build specific network architectures [30]. Indeed, while elegant work has been carried out elucidating the subtle factors involved in determining the molecular structures adopted by *s*-block metal complexes, no sustained attempt has been made to rationalize polymer formation in this area.

### 3.1 Common Geometries Available for Alkali Metal Containing SBUs

The use of lithiated SBUs was a reasonable starting point for the study since the understanding of the aggregation behavior for these complexes is by far the most developed within the *s*-block series [14-16]. Our initial phase of the project utilized the abundance of information that is available for lithiated species to choose suitably robust molecular aggregates as building blocks. **Figure 3** displays the core structures of a selection of commonly found aggregation states for *s*-block complexes: ring dimers, ring trimers, tetrameric cubanes, and prismatic hexamers. **Figure 3** also illustrates the relationship between the metal arrangements within the aggregates and their corresponding nodal geometry. The exact arrangement of the metals within each type of aggregate will vary with the identity of its anionic partner. This will lead to a variety of geometric distortions from the 'ideal' nodal configuration. Nevertheless, the basic ring or cage arrangements should remain intact and will therefore control the direction of polymer assembly.

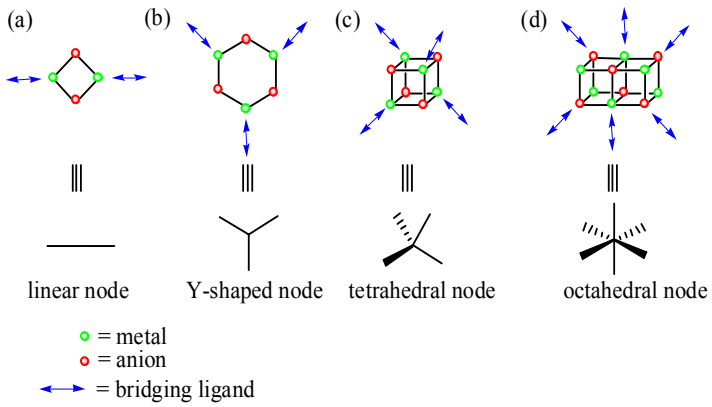



**Figure 3.** Relationship between selected metal aggregates and their related nodal geometries: (a) disolvated ring dimer, (b) trisolvated ring trimer, (c) tetrasolvated tetrameric cubane, and (d) hexasolvated prismatic hexamer.

### 3.2 Lithium Aryloxides as SBUs

It was essential to the success of our approach that the aggregation state of the metallated compound was predictable and remains intact both in solution and in the solid-state while in the presence of the Lewis base linker. An excellent starting point was lithium aryloxides. These complexes contain strong Li-O bonding, and numerous aggregate types are available as SBUs. In addition, they are readily prepared from the parent phenol by direct deprotonation using bases such as organolithium or lithium amide reagents. The aggregate size and coordination environment around the metal center can readily be controlled by the choice of phenol and donor solvent. For example, increasing the steric bulk at the 2,6-positions of an aryloxy ring results in decreasing the aggregation state of the lithiated complex. This is conveniently demonstrated for a series of THF-solvated lithiated aryloxides (ArOLi), shown in **Figure 4**, which form stable ring dimers (Ar = 2,6-$^{t}Bu_2C_6H_3$), ring trimers (Ar = 2,6-$^{i}Pr_2C_6H_3$), tetrameric cubanes (Ar = 2,4,6-$Me_3C_6H_2$), and prismatic hexamers (Ar = $C_6H_5$) [31-33].

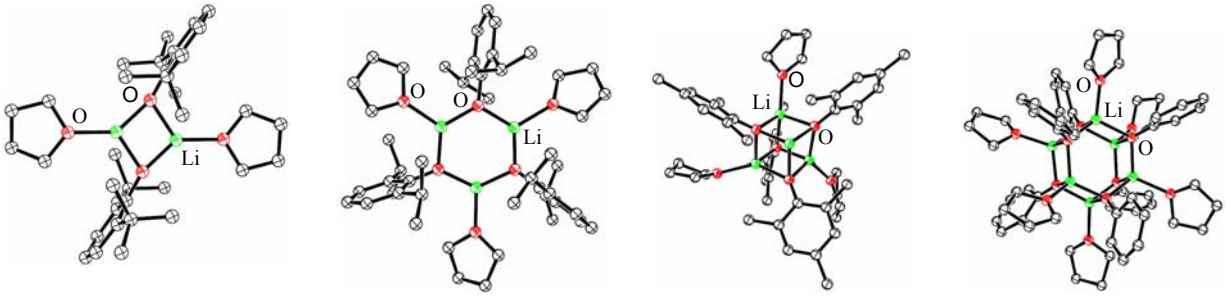



**Figure 4.** Molecular structures of a series of THF-solvated lithium aryloxides, $[ArOLi.THF]_n$.

Each of the metal atoms within these aggregates has one coordination site occupied by a monodentate THF donor molecule. This makes these anionic aryloxide ligands excellent targets as SBUs since replacement of the terminal Lewis bases by didentate linear-linker molecules will result in interconnecting aggregates. There are of course numerous possibilities for both the connectivity and topology of the supramolecular assemblies that may be formed via the interaction of the four aggregate types shown above with a linear-linker such as dioxane. However, it is becoming clear that a relatively few, high-symmetry structures dominate network topologies [26]. The most likely assemblies to be produced in each case are one-dimensional chains from disolvated dimers; hexagonal (6,3) sheets from ring trimers; three-dimensional diamondoid networks using tetrameric cubanes; and cubic lattices using prismatic hexamers.

### 3.3 Choice of Neutral Linker Ligand

Polydentate Lewis bases are required in order to connect neighboring aggregates. The strategy used to achieve this goal was to employ neutral linear-linker ligands containing relatively hard Lewis base donors [34]. A selection of potentially useful linkers for the *s*-block systems are shown in **Figure 5**.

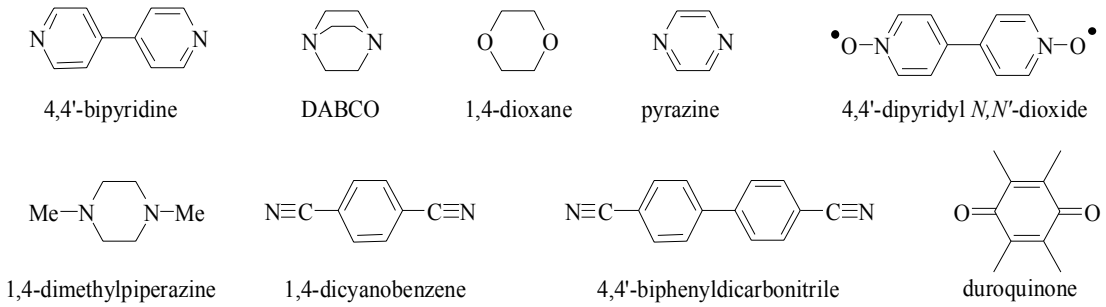



**Figure 5.** Neutral linear-linker molecules suitable for coordination to hard *s*-block metals.

Generally, divergent ligands have been used in lithiated systems simply to enhance the crystallization properties of a metallated molecular substrate in order to study localized bonding or metrical information, rather than to

deliberately build specific network structures [30, 35, 36]. However, in most instances no recognizable lithiated aggregate is present, rather isolated ions with no inherent directional control are linked together. Furthermore, reports on the use of other linear-linking didentate Lewis bases for the *s*-block metals are surprisingly scarce. This deficiency is remarkable considering the enormous importance of such ligands in the assembly of transition metal-containing networks [37]. Nevertheless, ligands of this type should be ideal candidates for coordination to *s*-block metals, in particular lithium, due to their excellent properties as hard Lewis bases. The monodentate base pyridine is a commonly used donor solvent for lithiated complexes with dozens of structurally characterized examples in the Cambridge Structural Database [38], and furthermore several crystal structures of lithium species have been characterized containing the chelating ligand 2,2'-bipyridine [39]. In the present study the linker of choice is 1,4-dioxane as this ligand proved to be convenient as it not only acts as a divergent linker but also as solvent media.

## 4. RESULTS AND DISCUSSION

### 4.1 Use of Tetrameric $Li_4O_4$ Cubanes as SBUs Linked by Dioxane

This early success in rationally designing complex network architectures from very simple constituents spurred us to expand the scope of our studies to include the possibility of linking together neutral, pre-assembled lithiated SBUs using *external* neutral divergent Lewis bases. This is appealing since it immediately opens up a wide variety of well-studied solvated aggregate types for investigation. We targeted the commonly encountered $Li_4O_4$ tetrameric cubanes formed by lithium aryloxides (ArOLi) for use as SBUs [31]. These complexes are excellent SBU candidates since they contain strong Li-O bonding, the metals are held in an approximately tetrahedral arrangement with one coordination site available for ligation, and there is a wide range of substituted phenols available for systematic studies. The didentate donor dioxane was selected as the divergent external linking Lewis base due to its capacity to act as solvent media, its rigidity and its inability to chelate a lithium center.

At the outset, an exploratory computational study was conducted to determine the feasibility of dioxane acting as a bridging ligand between a pair of $Li_4O_4$ cubanes, since this structural pattern had not previously been characterized (**Figure 6**). Semi-empirical PM3 calculations were the first line of analysis due to their computational efficiency [40, 41]. Geometry optimization calculations were performed on the full molecules [(PhOLi.dioxane)$_4$] and [{(PhOLi)$_4$}$_2$.(dioxane)$_7$] rather than using simplified model complexes to ensure an accurate portrayal of subtle steric and electronic effects. These calculations indicated that no notable energy penalty (<0.2 kcal/mol) is incurred on bridging a pair of tetrasolvated cubanes by the donor. This key result was confirmed by further geometry optimization at the *ab initio* HF/6-31G* level of theory (<0.01 kcal/mol) and also by high level single-point density functional theory calculations (B3LYP/6-311G**, <0.04 kcal/mol) [41-44]. These calculations firmly established the viability of polymer formation. In addition, this study demonstrates the ability of lower level calculations to be used as an efficient method for screening suitable aggregate/linker combinations.

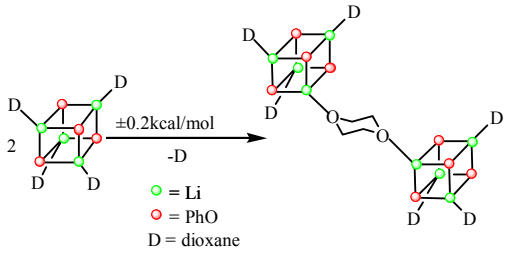



**Figure 6.** Calculated energetics of dioxane bridging a pair of $Li_4O_4$ units.

Appropriate lithium aryloxides were then chosen based on their predilection to form tetrasolvated $Li_4O_4$ cubanes in the presence of monodentate Lewis bases [23]. The required tetrameric aggregation was confirmed by preparing and structurally characterizing (by X-ray diffraction, XRD) the molecular analogues using monodentate donors (THF or pyridine), and also by solution NMR studies of the complexes in dioxane solvent media using the Jackman method of correlating the $^{13}C$ NMR shift positions for the *para*-carbons with aggregation state [25, 45]. Specific structural variants of these ligands were then selected as our understanding of these systems evolved. **Figure 7** highlights the structural characterization, by single-crystal XRD, of the three representative complexes [{(ROLi)$_4$.(dioxane)$_x$}$_\infty$], (a) R=Ph, x=3; (b) R=4-Et-$C_6H_4$, x=2.5; and (c) R=1-naphth, x=2, which typify our findings [46].

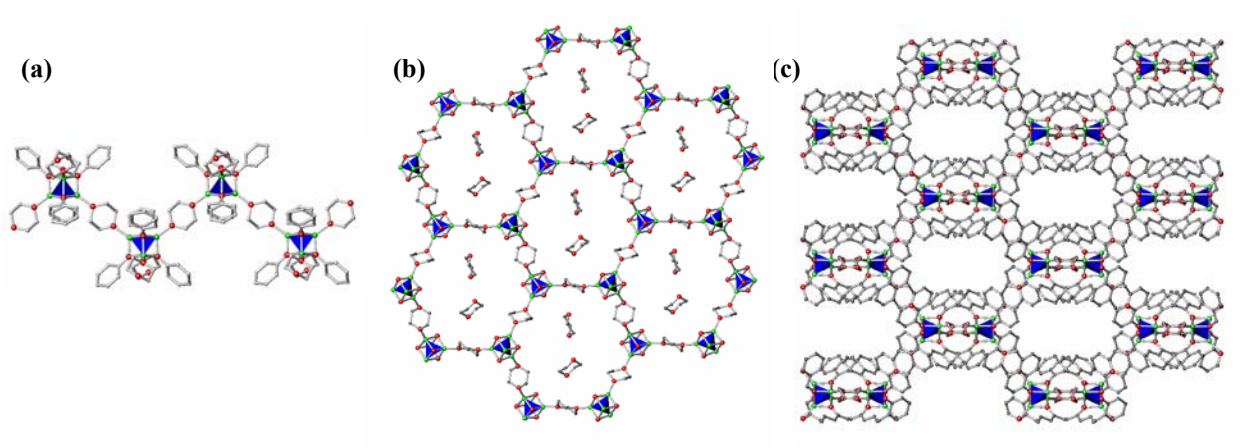



**Figure 7.** Sections of the polymeric structures of the: **(a)** 1D zig-zag chain, **(b)** framework atoms of the 2D hexagonal network with two guest solvent molecules per macrocycle, and **(c)** 3D diamondoid network. Vertices of the blue tetrahedra represent the four lithium centers within each cubane.

The first key point is that each structure is composed of linked $Li_4O_4$ tetrameric cubane units, proving that the desired molecular SBU remains intact on polymerization. Next, the three polymeric types found, 1D zig-zag chains, 2D hexagonal nets, and 3D diamondoid networks are very closely related, simply differing by the number of points of extension from the tetrahedral SBU (two, three and four leading to V-shaped, trigonal and tetrahedral nodes respectively). Moreover we have now fully characterized an extended series of $Li_4O_4$ substituted aryloxide complexes and found that they each form one of these three structural types. Pleasingly, this demonstrates that the SBU approach is indeed successful in limiting the number of network topologies for a given node/linker combination.

The complexes in **Figure 7** illustrate the dramatic effect of relatively small changes in sterics on the macromolecular structure adopted. We deduced that the type of polymer formed can be rationalized in terms of balancing entropic factors with the efficient filling of space. Although our calculations indicate that bridging by dioxane is thermoneutral, there will be a substantial increase in entropy associated with polymer formation due to the gain in translation freedom of the liberated solvent molecules. Next, there will clearly be an optimum length for the rigid anionic ligands of the SBUs in order to efficiently fill the hexameric macrocycles of the 2D sheets. If the ligands are too short this will result in energetically unfavorable large voids at the center of the rings. Conversely, if the ligands are too long the 2D arrangement will be destabilized by transannular interactions [47]. This analysis is consistent with the 1D chain structure observed for the small $PhO^-$ anions. Extension of the ligand at the *para*-position partially circumvents this problem and the longer 4-Et-$C_6H_4O^-$ anions, in combination with the guest dioxanes, fill the hexameric macrocycles with reasonable efficiency. However, examination of the extended crystal-packing diagram of this complex shows interdigitation between neighboring 2D layers, suggesting that significant space still remains within the sheet. In order for a diamondoid structure to be formed it is necessary to occupy 3D adamantanoid cavities. This proves possible for 1-naphthOLi by lateral extension to the aromatic ring, leading to excellent localized space filling in the 'corners' of the adamantanoid cavities by interlocking of the naphthyl units [47]. Space filling by interpenetration is precluded due to the $(1\text{-naphthOLi})_4$ cubanes being ~14.2 Å in diameter whereas the adamantanoid cavities have dimensions of only 7.3 x 7.6 x 16.6 Å. The open structure found for this complex supports the conjecture that the use of SBUs promotes the formation of open-framework materials. The viability and scope of our rationalization was tested through targeting the structural characterization of specific substituted aryloxides. This proved successful, such that when the *para*-substituent is small (H, Me, F or Cl) 1D chains are produced, minimal extension at this position (Et, OMe, $^i$Pr) leads to the formation of 2D hexagonal sheets, whereas lateral substitution of the aromatic (Ar = 1-naphth or 2,4,6-$Me_3C_6H_2$) gives 3D diamondoid networks. In all cases the homogeneity of the samples was confirmed by powder XRD of the bulk solids.

The potential for porosity is an exciting feature of the diamondoid structures outlined above and our preliminary investigations indicate that the guest molecules within $[\{(1\text{-naphthOLi})_4.(\text{dioxane})_2\}_\infty]$ (**Figure 7c**) can be removed under reduced pressure without disruption to the framework. Solvent loss was monitored by $^1$H NMR spectroscopy through complete dissolution of samples in $d_6$-DMSO, followed by integration of the anion and dioxane signals.

Initially, three enclatherated solvent molecules per $Li_4O_4$ unit are present but after 24 hours of evacuation ($1.5 \times 10^{-3}$ Torr and 30°C) the non-framework dioxane is entirely removed. Calculations indicate that 34.8% of the total volume within this structure is potential solvent space.[77] Comparison of the calculated, as prepared, and evacuated powder XRD patterns show only minor variations, demonstrating that the integrity of the framework is maintained on solvent loss. In addition, the macroscopic morphology of the crystals is retained after evacuation (perfect octahedra) allowing determination of the unit cell parameters by single-crystal XRD, and pleasingly they match those of the original material (a full structure determination has not yet been completed due to the weak diffraction of the crystals). Furthermore, these crystals were found to be completely air-stable over several weeks. Unit cell checks confirmed that the crystallinity is retained and IR spectroscopy showed no detectable decomposition to lithium hydroxide.

### 4.2 Use of Hexameric $Na_6O_6$ Prismatic Cages as SBUs Linked by Dioxane

Many complexes of the heavier Group 1 metals are appealing as SBUs. Of the heavier alkali metals, the alkoxides and aryloxides of both sodium and potassium have been reasonably well studied [48, 49]. A potential drawback of moving from lithium to the heavier Group 1 metals is a significant weakening of the metal-ligand bond strength, which may in turn lead to increased aggregate lability and a decrease in the predictability of the nature of the SBU. To test the feasibility of using these metal aggregates we targeted solvated $Na_6O_6$ hexameric aggregates as potential sodium based SBUs since this structural type has been characterized in the solid state for molecular derivatives, *e.g.* $[(PhONa.THF)_6]$ [50]. We theorized that the triple stack of dimers (face-shared cubane) aggregates should act as octahedral SBUs and lead to the formation of 3D cubic networks. Building on our work on the lithiated systems, we utilized 4-substituted aryloxides as convenient probes to test the effect of sterics on the supramolecular arrangements obtained. The set of *para*-halide substituted derivatives are particularly useful in illustrating the effect of sterics on network structure. The complexes $[(4\text{-R-}C_6H_4ONa)_6.(dioxane)_x]_\infty$, where (a) R = I, x = 5 form 1D linear chains, (b) R = Br, x = 4 from (4,4) 2D square nets, and (c) R = F and x = 3 form 3D cubic lattices, **Figure 8** [51]. This remarkable set of complexes again demonstrates that the desired hexameric $Na_6O_6$ SBU is robust and remains intact within each network. Also, the type of network can again be rationalized in terms of the size of the substituent at the *para* position, with only the fluorine group being small enough to fit into the available volume within the cubic cavities. Moreover, we have again confirmed that network formation using the $Na_6O_6$ SBUs is limited to these three structural types, *i.e.* structural characterization of the dioxane solvated *para*-aryloxides R-$C_6H_4ONa$: R = Cl, I or Ph, give 1D linear chains, R = Br, H, Me, $^i$Pr, $^t$Bu or $^t$Pn give 2D (4,4) nets and R = F or Et give 3D cubic networks. These results show that even the heavier *s*-block elements may be used as robust SBUs.

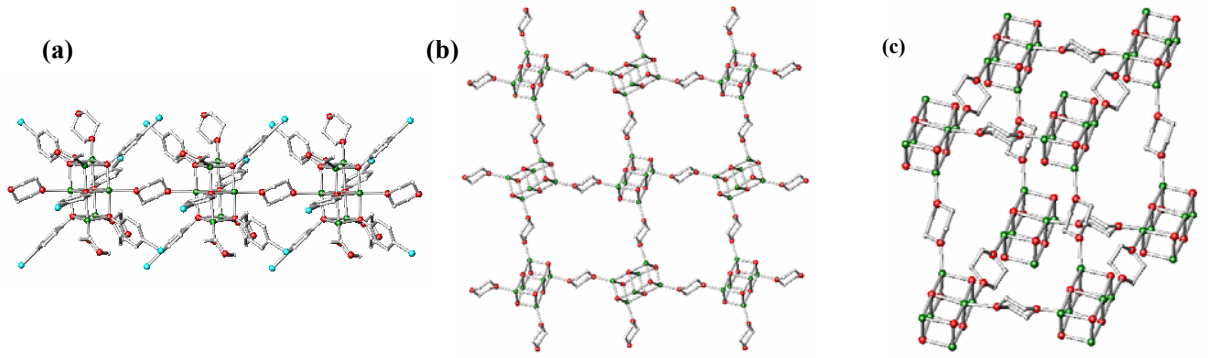



**Figure 8.** Sections of the polymeric structures of the: (a) 1D linear chain, **(b)** 2D square net (framework atoms only), and **(c)** 3D cubic lattice (framework atoms showing a single cubic unit), formed from triple-stack $Na_6O_6$ SBUs bridged by dioxane.

## 5. CONCLUSIONS

To summarize, we have demonstrated that both lithiated and sodiated pre-assembled molecular aggregates may be used to construct network assemblies. The topologies of the resulting frameworks are directed by the geometry of the metals within the molecular SBUs and is also dependant on the ability of organic components to fill space effectively. These initial studies suggest that there is a rich chemistry surrounding the use of early main group metals in the supramolecular synthesis of extended frameworks.

## 6. ACKNOWLEDGEMENTS

We gratefully acknowledge the Petroleum Research Fund (41716-AC3) and the University of Notre Dame for support. We also thank the National Science Foundation for instrumentation support (CHE-0443233).

## 7. REFERENCES

[1] L. Brammer, *Chem. Soc. Rev.* **33** (2004) 476.
[2] D. Braga, *Chem. Commun.* (2003) 2751.
[3] S. L. James, *Chem. Soc. Rev.* **32** (2003) 276.
[4] C. Janiak, *Dalton Trans.* (2003) 2781.
[5] J. K. M. Sanders, *Chem. Eur. J.* **4** (1998) 1378.
[6] G. B. Gardner, D. Venkataraman, J. S. Moore and S. Lee, *Nature* **374** (1995) 792.
[7] S. J. Lee and W. Lin, *J. Am. Chem. Soc.* **124** (2002) 4554.
[8] H. O. Stumpf, L. Ouahab, Y. Pei, D. Grandjean and O. Kahn, *Science* **261 (**1993) 447.
[9] J. W. Steed and J. L. Atwood, *Supramolecular Chemistry*, (Wiley, New York, 2000).
[10] G. R. Desiraju, *Crystal Engineering. The Design of Organic Solids*, (Elsevier, Amsterdam, 1989).
[11] I. Haiduc and F. T. Edelmann, *Supramolecular Organometallic Chemistry*, (Wiley-VCH, New York, 1999).
[12] V. Chandrasekhar, S. Nagendran, S. Bansal, A. W. Cordes and A. Vij, *Organometallics* **21** (2002) 3297.
[13] O. M. Yaghi, M. O'Keeffe, N. W. Ockwig, H. K. Chae, M. Eddaoudi and J. Kim, *Nature* **423** (2003) 705.
[14] T. Stey and D. Stalke, D. in: *The Chemistry of Organolithium Compounds*, eds. Z. Rappoport and S. Patai (Wiley, New York, 2004), p 47.
[15] A. M. Sapse and P. v. R. Schleyer, *Lithium Chemistry, A Theoretical and Experimental Overview*, (Wiley, New York, 1995).
[16] E. Weiss, *Angew Chem. Int. Ed. Engl.* **32** (1993) 1501.
[17] An excellent special addition of the Journal of Solid-State Chemistry is available: O. M. Yaghi, M. O'Keefe, eds. *J. Solid State Chem.* **152** (2000) 1-321.
[18] K. W. Henderson, A. R. Kennedy, A. E. McKeown and D. S. Strachan, *J. Chem. Soc., Dalton Trans.* (2000) 4348.
[19] S. E. Denmark, K. A. Swiss and S. R. Wilson, *Angew. Chem. Int. Ed. Eng.* **35** (1996) 2515.
[20] A. F. Wells, *Three-Dimensional Nets and Polyhedra*, (Wiley-Interscience, New York, 1977).
[21] J. Kim, B. Chen, T. M. Reineke, H. Li, M. Eddaoudi, D. B. Moler, M. O'Keeffe and O. M. Yaghi, *J. Am. Chem. Soc*. **123 (**2001) 8239.
[22] G. Boche, *Angew. Chem. Int. Ed. Engl.* **28** (1989) 277.
[23] K. W. Henderson, A. R. Kennedy, L. Macdonald and D. J. MacDougall, *Inorg. Chem.* **42** (2003) 2839.
[24] K. W. Henderson, A. R. Kennedy and D. J. MacDougall, *Inorg. Chem.* **42** (2003) 2736.
[25] K. W. Henderson, A. R. Kennedy and D. J. MacDougall, *Phosphorous, Sulfur and Silicon* **179** (2004) 795.
[26] N. W. Ockwig, O. Delgado-Friedrichs, M. O'Keeffe and O. M. Yaghi, *Acc. Chem. Res.* **38** (2005) 176.
[27] D. Braga, F. Grepioni and G. R. Desiraju, *Chem. Rev.* **98** (1998) 1375.
[28] B. Walfort, L. Lameyer, W. Weiss, R. Herbst-Irmer, R. Bertermann, J. Rocha and D. Stalke, *Chem. Eur. J.* **7** (2001) 1417.
[29] R. E. Dinnebier, U. Behrens and O. Falk, *J. Am. Chem. Soc.* **120** (1998) 1430.
[30] For a specific example of 1,4-dioxane bridging in a lithiated coordination polymer see: O. Kühl, J. Sieler, G. Baum and E. M. Hey-Hawkins, *Z. Anorg. Allg. Chem.* **626** (2000) 605.
[31] T. J. Boyle, D. M. Pedrotty, T. M. Alam, S. C. Vick and M. A. Rodriguez, *Inorg. Chem*. **39** (2000) 5133.
[32] K. Thiele, H. Gorls and W. Seidel, *Z. Anorg. Allg. Chem*. **624** (1998) 1391.
[33] L. M. Jackman, D. Cizmeciyan, P. G. Williard and M. A. Nichols, *J. Am. Chem. Soc.* **115** (1993) 6262.
[34] P. K. Chattaraj, H. Lee and R. G. Parr, *J. Am. Chem. Soc.* **113** (1991) 1855.
[35] F. T. Deelmann, F. Pauer, M. Wedler and D. Stalke, *Inorg. Chem.* **31** (1992) 4143.
[36] M. Motevalli, D. Shah and A. C. Sullivan, *J. Chem. Soc, Dalton Trans*. (1993) 2849.
[37] R. Robson, in: *Comprehensive Supramolecular Chemistry*, eds. J. L. Atwood, J. E. D. Davies, D. D. MacNicol, F. Vögtle and K. S. Suslick, (Elsevier, Oxford, 1996).
[38] F. H. Allen, *Acta Crystallogr.* **B58 (**2002) 380.
[39] For example see: B. W. Skelton, C. R. Whitaker and A. H. White, *Aust. J. Chem.* **43** (1990) 755.

[40] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, C. G. Zakrzewski, J. A. Montgomery Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman [illegible]7[illegible]cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, [illegible] Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, J. L. Andres, M. Head-Gordon, E. S. Replogle and J. A. Pople, *Gaussian 98*, revision A.6; Gaussian, Inc.: Pittsburgh, PA, 1998.
[41] W. J. Hehre, R. Ditchfield and J. A. Pople, *J. Chem. Phys.* **56** (1972) 2257.
[42] P. C. Hariharan and J. A. Pople, *Theo. Chem. Acta* **28** (1973) 213.
[43] A. D. Becke, *J. Chem. Phys.* **98** (1993) 5648.
[44] A. D. McLean and G. S. Chandler, *J. Chem. Phys.* **72** (1980) 5639.
[45] L. M. Jackman and B. D. Smith, *J. Am. Chem. Soc.* **110** (1988) 3829.
[46] D. J. MacDougall, J. J. Morris, B. C. Noll and K. W. Henderson, *Chem. Commun.* (2004) 456.
[47] S. Mecozzi and J. Rebek Jr., *Chem. Eur. J.* **4** (1998) 1016.
[48] R. E. Dinnebier, M. Pink, J. Sieler and P. W. Stephens, *Inorg. Chem.* **36** (1997) 3398.
[49] T. J. Boyle, N. L. Andrews, M. A. Rodriguez, C. Campana and T. Yiu, *Inorg. Chem.* **42** (2003) 5357.
[50] M. Kunert, E. Dinjus, M. Nauck and J. Sieler, *Chem. Ber.* **130** (1997) 1461.
[51] D. J. MacDougall, B. C. Noll and K. W. Henderson, *Inorg. Chem.* **44** (2005) 1181.

# MODELING STUDIES OF ANTI-AMYLOID ANTIBODIES BOUND TO Aβ

Anna S. Gardberg, Nathan Tanner, Chris Dealwis*
*University of Tennessee at Knoxville, Department of Biochemistry and Molecular Biology, M407 Walters Life Sciences*
*Knoxville, Tennessee 37996-0830*

Israel Huff, Kimberly Salone, Ronald Wetzel
*University of Tennessee at Knoxville, Graduate School of Medicine, Knoxville, Tennessee, USA*
*Corresponding author: cdealwis@utk.edu
Keywords: A-beta, beta-amyloid, monoclonal antibodies, Alzheimer's, neurodegenerative disorders Autodock, protein-protein interaction
Grant sponsors: AG, NT, IH, RW, and CD are supported by NIH grant 5R01NS046356-03. KS was supported by a grant from the American Health Assistance Foundation.

## 1. ABSTRACT

WO1 and WO2 are conformation-specific monoclonal IgMs that bind the fibril state of the amyloid Aβ peptide (1-40), as well as amyloid fibrils of other disease-related proteins. Significantly, these antibodies (Abs) do not bind the soluble, monomeric state of Aβ (1-40) or the precursor form of other amyloids. The Abs have been sequenced and compared with sequences in the Kabat database; though some unusual charged residues were noted in WO1 and WO2, none are at the CDR. Three-dimensional models of the Fv fragments of WO1 and WO2 were generated with Web Antibody Modeling. A trigonal prism and a two-strand structural model of the Aβ amyloid core were compared by docking each with the Fv models of WO1 and WO2. The results predict binding of WO1 and WO2 to the Aβ trigonal prism at the protofilament face consisting of residues 23-27 and of WO2 to the Aβ two-strand at the end of the elongating fibril. No positive prediction for WO1 and the Aβ two-strand was made. Binding experiments with WO1 were determined to be sensitive to salt and pH conditions, suggesting the importance of electrostatic interactions for binding; this result is consistent with the proposed docking of WO1 to the Aβ trigonal prism.

## 2. INTRODUCTION

The aggregation of normally soluble proteins into insoluble, unbranched fibrils is the underlying pathology of a family of diseases known as the amyloidoses.[1] The hallmark event in amyloidogenesis is a change in the secondary and/or tertiary structure of a normal, soluble protein, rendering it prone to self-assembly into highly ordered para-crystalline arrays: fibrils. More than 20 proteins have been clinically identified as precursors of amyloid fibrils *in vivo*. These include the amyloid precursor protein (APP), Islet amyloid polypeptide (IAPP), α-synuclein, transthyretin (TTR), immunoglobulin light chain (LC), polyglutamine-repeats, and prion proteins, that are associated with diseases such as Alzheimer's,[2-4] type II diabetes,[5,6] Parkinson's disease,[7-9] familial polyneuropathy,[10] light chain associated (AL) amyloidosis,[11-13] Huntington's disease,[4,14] and the spongiform encephalopathies.[15] By understanding the three-dimensional structure of such fibrils, we might design therapeutic agents to target them. It is notoriously difficult to extract structural information directly from amyloid fibrils, which are insoluble and non-crystalline, so a great deal of study has gone into fiber diffraction studies,[16-18] mutation studies,[19] microscopy,[17,20,21] proteolysis,[22,23] ESR,[24,25] NMR,[24,26] SANS,[27] and deuterium exchange with mass spectroscopy[28-31] in an ongoing attempt to elucidate the molecular structure of the Aβ fibrils associated with Alzheimer's disease.

There is no consensus model for Aβ structure, but most models incorporate cross-beta secondary structure. The cross-beta structure (in which the fibril axis is perpendicular to the chain direction) of Aβ models is derived from the 4.75 Å meridional reflection in fiber diffraction studies.[18] A stacked, parallel beta-sheet arrangement is suggested by Burkoth, Benzinger, and others,[24,27,32] while a pair of concentric cylinders was proposed by Perutz, *et al*.[33] An anti-parallel or stacked-hairpin beta-sheet arrangement is suggested by Callaway and others.[34-37]

One recent model of the Aβ protofilament, proposed by Guo, *et al,*[38] features a trigonal prism of stacked, parallel Aβ(15-36) polypeptides; the model is shown in **Figure 1a**. This trigonal prism protofilament model is consistent with threading analysis[38] and proline-scanning mutagenesis,[19] and the proposed fibril model (of 6 protofilaments) appears to be consistent with electron microscopy. Recent data from hydrogen exchange NMR[39] and cysteine scanning mutagenesis[40] have called portions of this construction into question.

Another proposed configuration of the Aβ protofilament consists of two stacks of two parallel strands twisting around one another along the fibril axis; this model, shown in **Figure 1b**, is consistent with data that was available in 2002, especially that from solid-state NMR studies.[26] It is also in line with the recent crystal structure of the amyloid form of a seven-residue fragment from yeast protein Sup35[41], which supports a double-beta-sheet structure. Both the two-strand and trigonal prism Aβ protofilament constructs are consistent with specific bodies of experimental data, but are not completely determined by experimental data. Both models continue to evolve with the acquisition of new data.

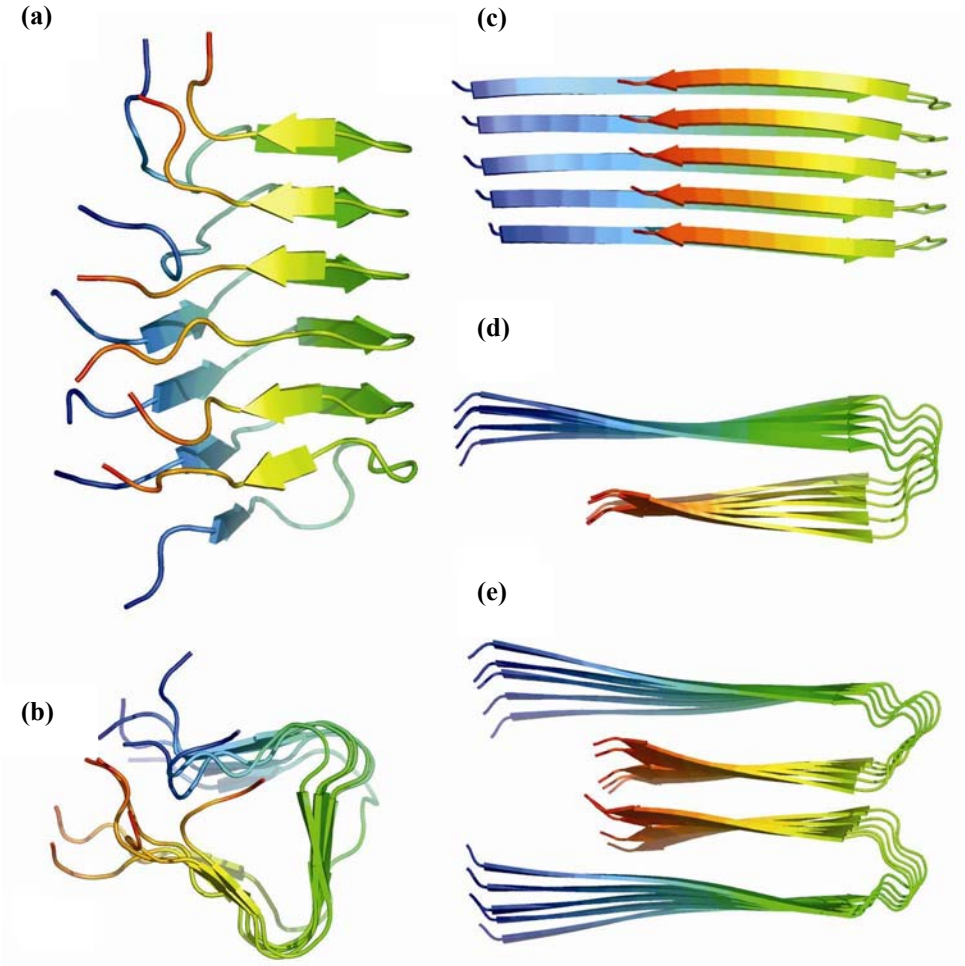



**Figure 1. (a)** Three-strand protofilament model in side view. **(b)** Three-strand, top view. **(c)** One half of proposed two-strand Aβ, side view. **(d)** Top view. **(e)** Two-fold symmetry about the *x*-axis has been used to generate one proposed configuration for the full protofilament from the two-strand model.

The monoclonal antibodies WO1 and WO2 (κ-light chain, murine, IgM) recognize a common conformational epitope shared by several different types of amyloid fibrils, with little dependence on amino acid sequence; the two bind the same antigen with similar affinities for the Aβ fibril.[42] This result opens the door to targeting whole classes of amyloid fibrils for detection and therapy.[43] Work to crystallize these antibodies (or their fragments) is ongoing. While IgM's are notoriously difficult to model,[44] the success of antibody modeling algorithms at predicting crystal structures of antibody variable regions[45] encouraged the use of computational models of WO1 and WO2 Fvs for docking simulations. Since WO1 and WO2 bind to Aβ amyloid, we hypothesized that a model of

the Aβ protofilament would, if correct, dock *in silico* to a model of the variable regions of WO1 and WO2. The configuration of the predicted complexes of the Fvs of antibodies WO1 and WO2 to two different models of Aβ protofilament is given. Experimental salt-dependence data is consistent with the prediction of WO1-Aβ binding for the trigonal prism model.

## 3. METHODS

### 3.1 Sequence Analysis

The nucleotide sequences of WO1 and WO2 were obtained by cloning and confirmed by multiple cloning using high fidelity polymerases, and further confirmed by extended N-terminal amino acid sequencing from analysis of the protein. Initial alignments were performed using Molecular Operating Environment (MOE™[46]), with Ab sequences obtained from the Protein Data Bank[47,48]. An extensive alignment was done using the Kabat sequence database testing program[49]. BLASTP analysis[50] was run on the WO1 and WO2 variable light chains using the Non-Redundant Protein Database at the San Diego Supercomputer Center. Canonical classes are based on Chothia nomenclature[51] and numbering scheme is that of Kabat[52]. The WO1 and WO2 Fv residue sequences were initially compared to one another. Then they were compared with sequences from the Kabat database to search for anomalies, unique features, or homologies to other antibodies. A sequence alignment of the CDRs is shown in Figure 2.

### 3.2 Model Building

Three-dimensional structures of WO1 and WO2 variable regions (Fv) were generated using the Web Antibody Modeling algorithm, WAM.[45] WAM is an improvement on the AbM program [53-56] with greater capability in modeling the highly variable H3 loop through a combination of knowledge-based and *ab initio* methods[45] . The WAM algorithm has a record of producing models that are 1.0-2.8 Å RMSD from observed structures for the heavy chain CDR3, and better than that for the canonical loops (typically 1.0-2.5 Å RMSD).[45]

### 3.3 Docking

#### 3.3.1 Preparation

A PDB file of the trigonal prism Aβ 1-40 hexameric protofilament model after molecular dynamics simulations ("Aβ trigonal prism", which models residues Gln15 through Val36)[38] was kindly provided by Juntao Guo and Ying Xu. A PDB file of the two-strand pentameric Aβ 1-40 protofilament model after energy minimization ("Aβ two-strand", which models residues 9 through 40)[26] was kindly provided by Robert Tycko.

The program Autodock3[57] predicts the interaction of ligands with macromolecular targets. As Autodock3 can handle no more than 2048 atoms, we reduced the two-strand protofilament to a tetramer and did not generate its symmetry pair. Control calculations performed with a protein-peptide complex, a protein-ligand complex, and a lysozyme-anti-lysozyme complex of known structure (PDB codes 1SMR, 1DQJ and 1DYI) showed that the charges assigned from forcefield calculations performed by AutoDockTools yielded less accurate results than those assigned by AMBER forcefield calculations[58,59] performed by InsightII.[60] We assigned charges using the following protocol. The atom names were standardized to Refmac5 conventions. Using Refmac5[61,62], hydrogen atoms were added in “riding” positions computed from the carbon, nitrogen, oxygen, and sulfur atomic positions. After capping the termini of each chain with (neutral) carboxylic and amine end groups, the partial charges for each atom (including polar hydrogen atoms) in each model were calculated with the InsightII program using the AMBER force field. Calculating the summed formal and partial charges for each model verified that the model is electrically neutral. A Sybyl-style file was then exported from InsightII.

#### 3.3.2 Computation

Autodock-style PDBQ files were prepared from the Sybyl-style files (described in the previous section) with *mol2topdbq* and *mol2topdbqs*, awk-based utilities packaged with Autodock3.[57] For purposes of solvation calculation, we designated the Fv models the “proteins” and the Aβ protofilament models the “ligands”. Despite the thermodynamic importance of side-chain motion,[63] neither the Fvs nor the Aβ models were permitted any torsional freedom; the Autodock package does not permit the side chains of the protein to move, and the Autotors utility for designating rotatable bonds in the ligand allows a maximum of 32, not nearly sufficient for a the 2395-2040-atom Aβ models. Grid parameter files were generated *via mkgpf3* (a script which comes packaged with

Autodock3) and adjusted to compensate for the large volume required for the interaction of the Fab's complementarity determining region (CDR) with the protofilament models. The grid maps were calculated using AutoGrid. The gridded area dimensions were 66 Å x 66 Å x 66 Å, with grid spacing increased to 0.55 Å. The final gridded regions encompassed the CDR of the Fv (as well as all of the heavy chain and most of the light chain of the model for WO1, and all save the C-termini for WO2) and the starting position of the Aβ model, with room for rotation and translation. Grid generation was performed with Autogrid3.
Docking was performed using the Lamarckian genetic algorithm (LGA), and the pseudo-Solis and Wets methods were applied for the local search. Each docking experiment was performed 50 times. Default parameters were applied for the docking experiments, with the exception of the initial rotation step size, which was set to 180. The results of the docking experiments were evaluated by calculating the positional root-mean-square (rms) deviation of the corresponding atoms of each conformation. Docking parameter files were prepared with the *mkdpf3* setup utility and adjusted by hand.

### 3.3.3 Contact Analysis

Contacts between the Fv models and the protofilament models in the predicted docking complexes were analyzed with the program CONTACT.[62] Docking clusters and configurations were examined Pymol,[64] with residue charge surfaces visualized. Electrostatic surface plots for publication were generated by the DelPhi module of InsightII.[60] Angles between solutions were computed by calculating the average direction vector along the protofilament axis for each solution and taking the inverse cosine of the dot-product of direction vectors.

### 3.4 Electrostatic Analysis

Electrostatic surfaces were calculated using DelPhi, a part of the InsightII graphical software package. The following parameters were employed: solvent dielectric 80, radius 1.4, ion radious 2.0, grid solvent extent 15, 65 points. Models were displayed and manipulated with InsightII and all computational work done on a Silicon Graphics computer system.

### 3.5 Salt Effects on Binding WO1

Binding of the WO1 antibody to Aβ fibrils was tested under 13 different conditions of cation, anion, salt concentration, and pH. 1xPBS buffer was used as a control. Experiments 1-5 comprised the salt-effects experiments; they consisted of: 5 mM HEPES at pH 7.5 with: 1) no salt, 2) 150 mM NaCl, 3) 600 mM NaCl, 4) 150 mM KCl, and 5) 75 mM $Na_2HPO_4$. In experiments 6-13, the effects of pH were examined; they consisted of:150 mM NaCl with 6) citric acid pH 3.0, 7) citric acid pH 5.8, 8) PIPES at pH 5.8, 9) PIPES at pH 7.4, 10) Bicine at pH 7.4, 11) Bicine at pH 8.5, 12) CHES at pH 8.5, and 13) CHES at pH 10.0.

No ionic interference was expected from HEPES, which has low ionic strength. $Na_2HPO_4$ has roughly double the ionic strength of the other two salts so it was used at half the concentration for a balanced comparison. The salt-effect experiments were buffered to pH 7.5, near physiological pH conditions. The pH effect experiments contained 150 mM NaCl to provide approximate physiological ionic strength conditions. The Aβ fibrils were prepared as described in *Kheterpal, et al, 2000*.[31]

## 4. RESULTS

### 4.1 Sequence Analysis

### 4.1.1 Comparison of WO1 to WO

The light chain variable regions share 66% identity (80% similarity) and the heavy chain variable regions share 82% identity (85% similarity). Overall charge composition of the Fv fragments is very similar, with isoelectric points (pIs) of the variable regions calculated to be: WO1 $V_H$ 7.3, $V_L$ 8.7, WO2 $V_H$ 7.2, and $V_L$ 8.7.

While both sequences are rich in hydrogen bonding residues, WO1 $V_L$ CDR has a greater proportion of charged Asp, Glu, Arg, and Lys residues (19% vs. 11%), while the WO2 $V_L$ CDR is richer in uncharged, polar hydroxyl residues (42% for WO1 $V_L$, 64% for WO2 $V_L$). The sequences of the CDR's are compared in **Figure 2**. The WO1 and WO2 heavy chain variable regions are more similar in overall sequence, and the number of charged residues in the $V_H$ is identical. Still, the two charged residues of WO1's $V_H$ are at its CDR, and those of WO2's $V_H$ are not at its CDR. WO1's light chain CDR contains charged residues at positions occupied by hydroxyl residues in WO2 such as Arg24, Lys52, and Asp56 in WO1 which are replaced by Thr, Ser, and Ser in WO2. Similarly, the unique charged

residue Arg93 in WO2 replaces a serine residue in WO1. WO1's greater proportion of charged residues and WO2's greater proportion of uncharged polar residues suggest that WO1 and WO2 bind amyloid differently from one another.

a)

| VL CDR1 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 31a | 32 | 33 | 34 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| WO1 | R | A | S | G | N | I | H | N | - | ***Y*** | L | A |
| WO2 | T | A | S | S | S | V | S | S | S | ***Y*** | L | H |
| α-Lyso | R | A | S | Q | N | I | S,H | N | N | - | L | H,A |
| Germ. | R,K | S,A | S | Q | S,N | I | nh* | nh | S | nh | L | nh |
| Kappa | R,K | A,T | S | Q | nh | V,I | nh | S,T | - | nh | L | nh |

| CDR2,3 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| WO1 | N | A | K | T | L | ***A*** | D | Q | H | F | W | S | T | P | ***Y*** | T |
| WO2 | S | T | S | N | L | ***A*** | S | L | Q | Y | H | R | S | P | ***Y*** | T |
| α-Lyso | Y | T | T,S | Q,T | S,L | nh | D,S | Q | Q | F,Y | nh | S | - | P | R,L | T |
| Germ | nh | A,V | S,A | nh | R,L | D,E | S | nh | Q | nh | S | nh | nh | P | R | T |
| Kappa | nh | A | S | nh | R,L | nh | S | Q | Q | S,Y | nh | S,D | nh | P | R,L | T |

b)

| VH CDR1,2 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| WO1 | G | Y | T | F | T | E | Y | ***T*** | M | H | G | I | ***N*** | P | N | ***N*** | G | T | **S** |
| WO2 | G | Y | S | F | T | G | Y | ***T*** | M | N | L | I | ***N*** | P | Y | ***N*** | G | T | **S** |
| α-Lyso | G | Y | T,S | F | T | T | Y | W | I,V | E,S | E | I | L | P | G | S | nh | T | Y,D |
| Germ | G | F | T | F | T | D | Y | Y | M | S,H | F | I | R | nh | nh | nh | nh | T | E,K |
| Kappa | G | F,Y | T | F | T | D | Y | Y | I,M | H,K | nh | I | nh | P | nh | nh | G,S | T | nh |

| CDR3 | 96 | 97 | 98 | 99 | 100 | 101 | 102 | 103 | 104 | 105 | 106 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| WO1 | C | A | R | D | D | G | - | ***Y*** | Y | G | F |
| WO2 | C | A | R | R | L | G | D | ***Y*** | Y | A | M |
| α-Lys | C | A | R,S | nh | D,E | G | - | nh | Y | G | nh |
| Germl | C | A | R,S | D | nh | nh | nh | nh | Y | A,Y | M,G |
| Kappa | C | A | nh | G,D | nh | G | nh | nh | nh | nh | F |

**Figure 2.** CDR composition of a) $V_L$s and b) $V_H$s compared with those of anti-lysozyme, germline, and kappa-light-chain antibodies. **nh* signifies no homology. Bold, italicized text highlights conserved residues in WO1 and WO2 White-on-black text indicates identical or homologous residues for WO1, WO2, and one or more compared sequences

### 4.1.2 Light Chain

A preliminary sequence alignment was performed using a database of various antibody sequences, including subsets of anti-lysozyme Ab chains and germline Ab chains. The results showed a high degree of homology between heavy and light chain sequences, with several residues noted as distinctive in **Table 1**. The CDRs of the light and heavy chains of WO1 and WO2 are rich in hydrogen-bonding residues relative to anti-lysozyme and germline antibodies, suggesting that hydrogen bonding plays an important role in amyloid-recognition.

A more extensive alignment was done using the Kabat sequence database testing program. The WO1 $V_H$ and $V_L$ sequences were compared to the database of 2707 light chains and 3471 heavy chains. The unique residues Lys42 and Lys74 had no consensus within the database; at these positions, which are not at the CDR, Lys occurred in less than 1% of the sequences. Lys52 at light chain CDR2, however, was a hydroxyl residue in the majority of the 2701 chains. None of the WO2-unique charged or hydroxyl residues showed any incongruity.

| Light Chain Residue WO1 WO2 | Equivalents | Residue exhibited by chains with non-identity |
|---|---|---|
| Lys42 Ser43 | No | None charged, only Gly, Gln, Tyr |
| Lys52 Ser53 | Lys only in WO1 | No charged, 95% Tyr or Ser |
| Asp56 Ser57 | Asp in WO1 and (AL) | 99% Pro, (G) have Thr/Ser |
| Lys74 Thr75 | Lys only inWO1 | 100% Glu/Asp, (G) has Thr |

**Table 1:** Light chain-unique residues and comparison for WO1 residues that are less than 10% homologous to database sequences in the initial alignment tests. The alignment was done with specific interest in the anti-lysozyme Abs as well as germline Abs, and any residues here are noted if appearing in either chain. The residue is listed with its location, as well as the residues that were most commonly shared throughout the other Ab sequences. The equivalent residue from WO2 is also included for comparison. Codes: antilysozyme (AL), germline (G).

### 4.1.3 Heavy Chain

Through the preliminary alignment including the anti-lysozyme, germline, and kappa-light chain sequences, several residues were noted as distinctive. They are listed in **Table 2**. As with the light chain results, the most striking disparities between the WO1/WO2 heavy chains and comparison germline and anti-lysozyme sequences occur with charged residues. Again, the Kabat alignment was performed. The equivalent positions for residues Lys63, Lys65, Lys67, Lys74, Asp73 (in WO1 and WO2) and Arg84 (in WO1 only) have no consensus, but these charged residues are not unique and do not occupy the CDR. The only significant discrepancy was the Asn44 of the WO2 heavy chain. The Asn residue found at this site is found in only 13 examples, 0.326% of the database. Typically, a Ser residue is at this position. This position is also not at the CDR.

| Heavy Chain Residue WO1 WO2 | Equivalents | Residue exhibited by chains with non-identity |
|---|---|---|
| Lys63 Lys63 | WO1/WO2, (G) | No other charged residues |
| Lys65 Lys65 | WO1/WO2, (G), (AL) | 100% Val, Leu, or Phe |
| Lys67 Lys67 | WO1/WO2, (G), (AL) | 100% Gly, Ser, Asp |
| Asp73 Asp73 | WO1/WO2, (G), (AL) | Mostly Asp/Glu |
| Lys74 Lys74 | WO1/WO2, (G) | 99% Asp/Glu |
| Arg84 Leu84 | Arg only in WO1 | Non-consensus, but no other charged |

**Table 2:** Heavy chain unique residues and comparison. The residues displaying less than 10% homology in the initial alignments to database sequences are included here. Any germline or anti-lysozyme Ab commonalities are noted, as well as the residues which occur in the chains which are not homologous to the WO1/WO2 heavy chains. Codes: antilysozyme (AL), germline (G).

### 4.1.4 Fv Models

Of the WO1 light chain's unique lysine residues (42, 52, and 74), only Lys52 is at the CDR and may play a role in docking. However WO1 light-chain Asp56 (unique to WO1 and anti-lysozyme Abs) may play a role, since it makes a hydrogen bond to an Asn residue in the trigonal prism model.

### 4.2 Electrostatic Surface Potentials for WO1 and WO2

WO1 and WO2 Fvs display regions of significant charge density (as calculated by the DelPhi module of InsightII), observed in the electrostatic potential map shown in Figure 3. Though the pIs of the WO1 and WO2 Fvs are nearly neutral, they clearly have highly charged regions. The presence and distribution of these charged segments indicates a possible role of charge interactions in antigen binding. The line of positively charged residues on each of the light-chains suggests a template for hydrogen bonding and salt bridges. The line of negatively charged residues at the heavy-light interface of WO1's CDR (as seen in **Figure 3**) reflects the higher proportion of negatively charged residues in its composition.

### 4.3 Electrostatic Surface Potentials for the Protofilament Models

The most striking features of both models are the bands of charged residues aligned along the stacking axis; these representations are shown in **Figure 4**. The charge bands are a consequence of the in-register parallel-stacking of the models. Labeling the trigonal prism protofilament model's face containing residues 17-22 "A", 23-27 "B", and 28-36 "C" (Aβ 1-40 numbering basis), we see that faces A and C display exposed hydrophobic residues. In the two-strand model, the N-terminal face consists of residues 9-23; this corresponds to the "A" face of the trigonal prism

model. There is then a 180º turn consisting of residues 24-27; this turn (instead of a β sheet) is in sharp contrast to the "B" face of the other model. Finally, residues 30-40 form the C-terminal face of the two-strand model.

**(a)** **(b)**

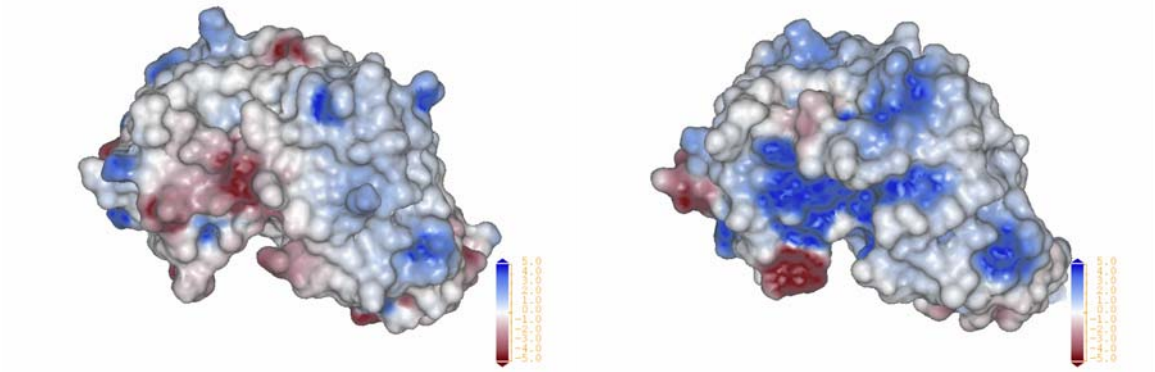


**Figure 3**. Electrostatic surfaces for WO1 **(a)** and WO2 **(b)** Fv models. The viewer faces the CDRs. In each image, the heavy chain fragment is on the left.

**(a)**

**(b)**

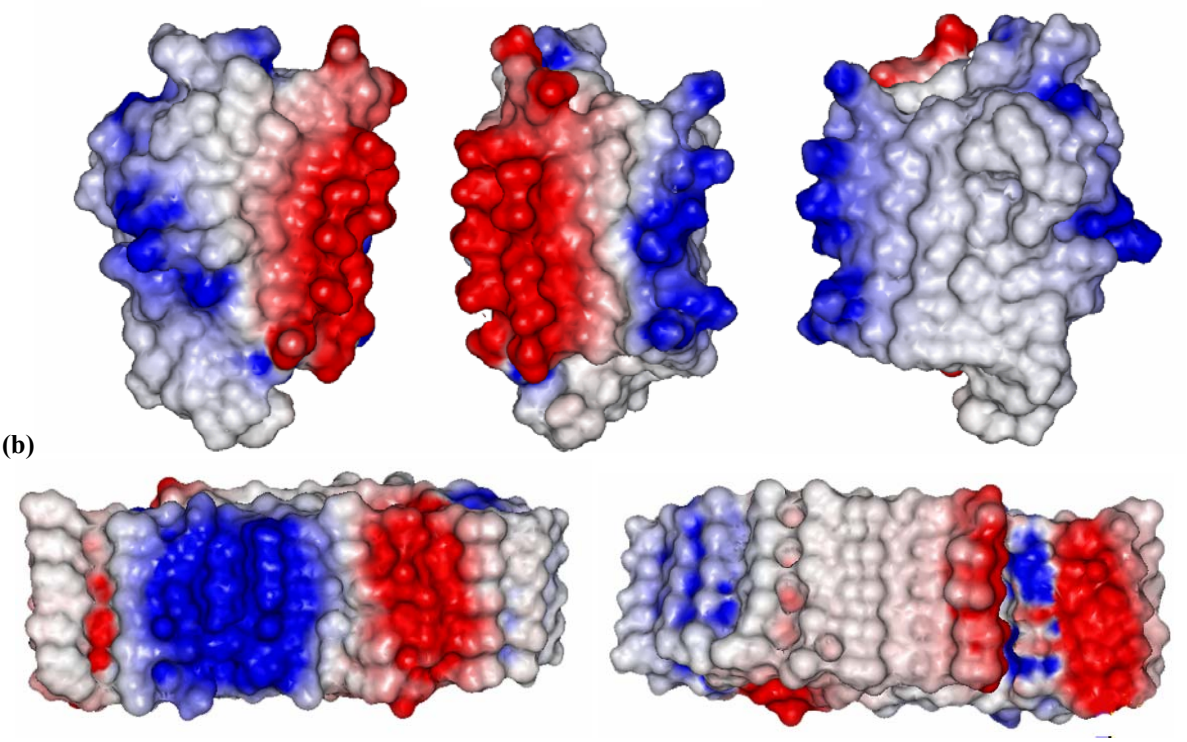


**Figure 4**. **(a)** Electrostatic surfaces for the A, B, and C faces of the trigonal prism protofilament model. The line of Glu22 and Asp23 residues (shown in red) at the corner of the A and B faces is particularly striking. **(b)** Electrostatic surfaces for the two faces of the two-strand model. The N-terminal face is shown in the left-hand image; strand direction is left to right. The C-terminal face is shown in the right-hand image; strand direction is left to right.

### 4.4 Docking WO1-Trigonal Prism-Model

AutoDock3 predicted 50 binding configurations. Both the Fv and the trigonal prism model were treated as rigid bodies. Autodock3 found 41 distinct conformational clusters (using an RMSD-tolerance of 1.0 Å), of which 8 contained more than one solution. The variation in energy from the best to the worst solution was 7.0 kcal/mol.

Although there was one cluster of two solutions in the top 7 (with the B face laterally at the heavy-light interface), more intriguing is a collection of related clusters beginning at the 8th-ranked solution. In total, 21 solutions have the B face of the trigonal prism protofilament model longitudinally oriented at the light-chain CDR, with some overlap onto heavy chain CDR. Some are offset by one, two, or three monomers (i.e., the stack of six Aβ monomers is shifted along the stacking axis), others by angles less than 27˚. This is the largest supercluster; its lowest-energy solution is presented in **Figure 5a**. The particular solution used for the following analysis was the lowest-energy member of this supercluster, which is 3.5 kcal/mol (~1 hydrogen bond) less favorable than the lowest-energy solution of all docking runs.

### 4.4.1 Salt Bridges and Ion Pairs

This docking model indicates that the line of Glu22 and Asp23 residues in the trigonal prism model binds to the corresponding line of polar and positively charged residues of WO1's light-chain CDR through salt bridges and type 2 hydrogen bonds between acidic residues and carbonyl groups. The line of Lys28 residues likewise interact with the trail of polar and negatively charged residues of the light *and* heavy chain CDRs. Breaking the ion pairs down by Szilyagi and Zavodsky classification,[65] there are four bonds at < 4.0 Å, six bonds between 4.0 and 6.0 Å, and ten pairs between 6.0 and 8.0 Å. *Hydrogen bonding*. The O and N atoms of the side chains of the line of Asn27 residues on the three-strand model make hydrogen bonds with Tyr and Asp residues of WO1. The carbonyl O atoms of the line of (inward-pointing) Val24 and Gly25 residues bond to Tyr residues of WO1. The backbone N of inward-pointing Ser26 makes some possible hydrogen-bonds. There are 17 likely hydrogen bonds at less than 3.5 Å, two of which were also counted as close ion pairs. The hydrogen bonding interface is shown in **Figure 5d**.

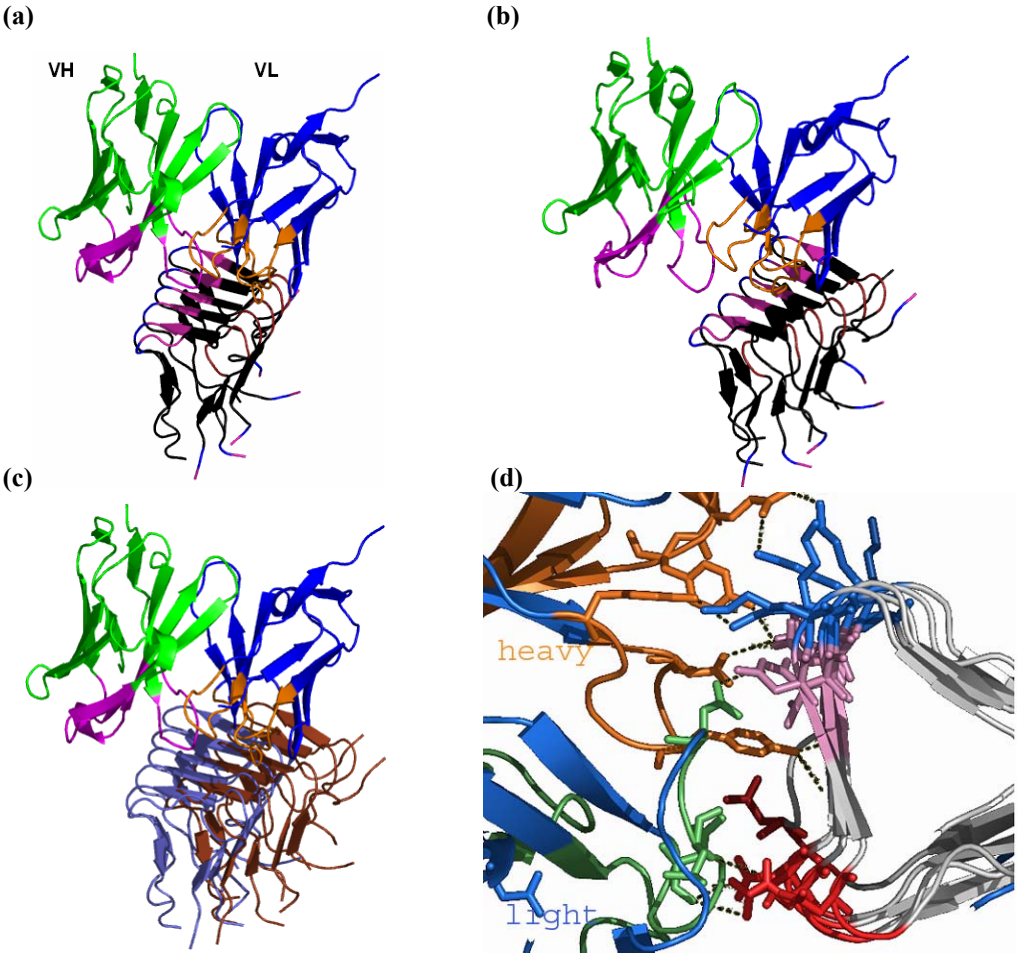



**Figure 5**. **(a)** WO1 and **(b)** WO2 Fv docked with trigonal prism protofilament model in the predicted configuration (top solution in supercluster). The heavy chain is on the left, the light chain is on the right, and the CDRs are highlighted. The trigonal prism model is shown colored according to residue: negative, red; positive, blue; polar, pink; hydrophobic, black. **(c)** The two solutions are showed overlain. **(d)** A close-up view of the docking interface for the WO1-trigonal-prism model. The color code is as for **(a)** and **(b)**, but hydrophobic residues are shown in grey and hydrogen bonds are shown in black.

**4.4.2 Hydrophobic Interactions**
There are no aromatic rings on the B-face of the trigonal prism model, and the hydrophobic Val24 side chain points toward the interior of the trigonal prism model. Thus hydrophobic interactions between WO1 and the trigonal prism model are not observed in this docking model.

**4.5 Docking WO2-Trigonal Prism-Model**
AutoDock3 predicted 50 binding solutions. The Fv and the trigonal prism model were both treated as rigid bodies. For WO2 and the trigonal prism protofilament model, Autodock3 found 42 distinct conformational clusters (using an RMSD-tolerance of 1.0 Å), of which 5 contained more than one member. The variation in energy from the best to the worst solution was 9.3 kcal/mol. There was one cluster of two solutions in the top 13, but the largest cluster began with the $14^{th}$-ranked solution. Within 1.0 Å R.M.S.D., there are 5 similar solutions. In total, 11 solutions have the B face of the trigonal protofilament model at the light-chain CDR, with slight overlap onto heavy chain CDR. One of them is offset by one monomer, others by acute angles. This is the largest supercluster; its lowest-energy solution is shown in **Figure 5b**. This solution resembles the supercluster solution found in the WO1 docking, but makes a 16° angle to it. The particular solution used for the following analysis was the lowest-energy member of this supercluster, which is 6.4 kcal/mol less favorable than the lowest-energy solution.

**4.5.1 Salt Bridges and Ion Pairs**
The analysis for this docked conformation is very similar to that of the WO1-Aβ6 solution, but there are no salt bridges at less than 4.0 Å, only two ion pairs between 4.0 and 6.0 Å, and five between 6.0 and 8.0 Å. *Hydrogen Bonding and hydrophobic interactions.* Thirteen likely hydrogen-bonds at less than 3.5 Å exist between WO2's Fv model and the docked trigonal-prism model. No hydrophobic interactions were observed for this docking model.

**4.6 Docking. WO1-Two-Strand-Model**
AutoDock3 predicted 50 binding solutions. Both the Fv and the protofilament model were treated as rigid bodies. For WO1 and the two-strand protofilament model, Autodock3 found 4 distinct conformational clusters with more than two members (using an RMSD-tolerance of 2.0 Å. The variation in energy from the best to the worst solution was 5.6 kcal/mol. In total, 4 solutions have the N-terminal face on the CDR, 4 have the open end of the two-strand model sitting longitudinally on the CDR, 7 solutions have the "bottom" end (right-hand-rule) of the two-strand protofilament model laterally across the CDR, and 3 have the closed end of the two-strand model near the CDR. The several clusters of near-equal energy suggest a negative result; they are shown in **Figure 6a**. Two of these clusters are not predicted to make any salt bridges at < 4 Å, but the other two are expected to make one and two such bonds, respectively.

**4.7 Docking. WO2-Two-Strand-Model**
As before, AutoDock3 predicted 50 binding solutions with both components (the Fv and the protofilament model) treated as rigid bodies. For WO2 and the two-strand protofilament model, Autodock3 found 35 distinct conformational clusters (using an RMSD-tolerance of 2.0 Å), of which 3 contained more than two members. The variation in energy from the best to the worst solution was 8.6 kcal/mol. The top 7 solutions have the "bottom" end (right-hand-rule) of the two-strand protofilament model laterally across the CDR. This cluster averages 3.8 kcal/mol lower in energy than the next best solution and it is the largest cluster, indicating a positive result; it is displayed in **Figure 6b**. This somewhat resembles the 7-member cluster for WO1-two-strand-model (shown in **Figure 6a**). The particular solution used for the following analysis was the lowest-energy member of the cluster.

**4.7.1 Salt Bridges, Ion Pairs, Hydrogen Bonding, and Hydrophobic Interactions**
No salt bridges were observed out to 4.0 Å, but 10 ion pairs in the range of 4-6 Å suggest a modest electrostatic contribution to WO2 binding Aβ. WO2 and the two-strand model make 7 likely hydrogen bonds at less than 3.4 Å. No *pi*-overlaps are observed.

**4.8 Docking Controls**
Positive controls #1 and #2 used rigid models, positive control #3 allowed flexibility in the ligand. Positive control #1 successfully recreated the crystal structure 1DQJ,[66] lysozyme complexed to antilysozyme. Two of the 50 solutions clustered near that from the crystal structure. These two averaged 11.8 kcal/mol lower in energy than the next most energetically favorable solution. Positive control #2, mouse renin complexed to the angiotensin analog CH-66 (PDB code 1SMR),[67] showed a similar result, with the 44-member dominant cluster 19.7 kcal/mol lower in energy. Positive control #3 did not precisely recreate the crystal structure 1DYI,[68] DHFR complexed with

folate, but rather predicted 10 similar solutions for the 32-atom folate molecule in the correct binding pocket within 2.3 to 10.8 Å RMSD of the true crystal structure, and within 3.5 kcal/mol of one another. Negative control #1 consisted of "docking" the trigonal prism protofilament model with the crystal structure of mouse renin. The result consisted of three superclusters (23, 14, and 5 members) of solutions whose lowest-energy members were within 5.0 kcal/mol of each other. Negative control #2 consisted of "docking" the WO1 Fv model with the inhibitor from the mouse renin crystal structure. All fifty top solutions were within 4 kcal/mol of one another. There was one two-member cluster and one four-member cluster. Solutions ranged all over the Fv.

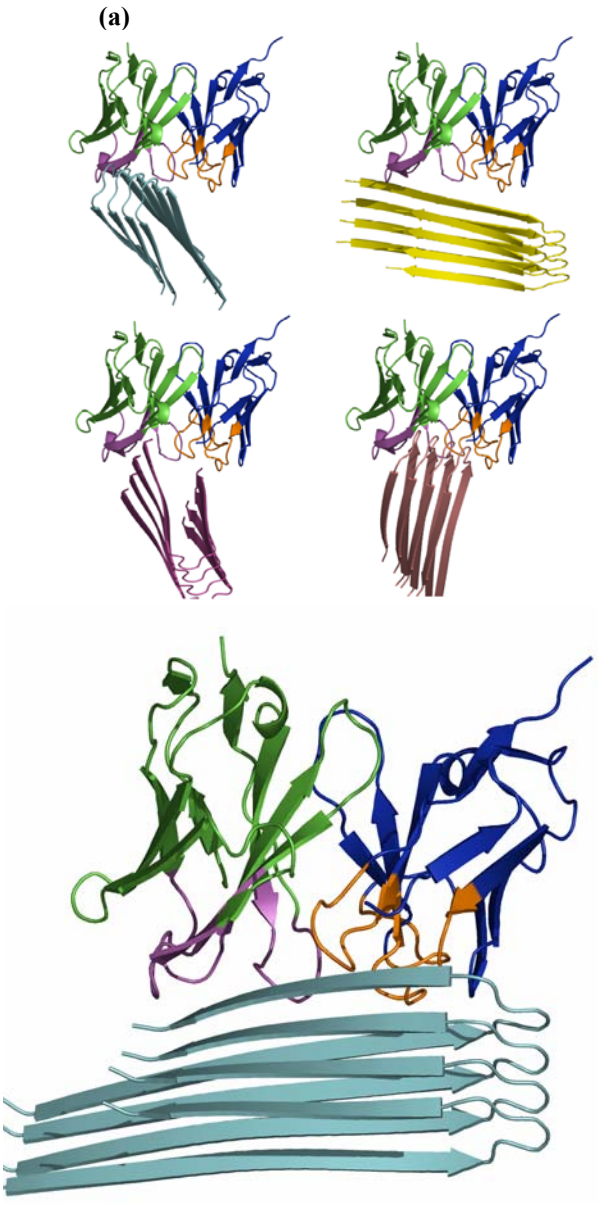



**Figure 6.** **(a)** WO1 Fv docked with two-strand protofilament model in the four predicted configurations. **(b)** WO2 Fv docked with two-strand protofilament model in the top predicted configuration

### 4.9 Binding vs. Salt Concentration

Binding constants were calculated from the sigmoid midpoints of the graph in **Figure 7**. The strongest binding of WO1 and Aβ was observed in PBS; the $EC_{50}$ was 2 nM. Of the 5 mM Hepes conditions, the ones at 150 mM NaCl (nearly-physiological) and at no salt resulted in the strongest binding of WO1 and Aβ, 4 nM. In the presence of 600

mM NaCl, the $EC_{50}$ constant changed sharply to 20 nM, indicating five-fold less binding. Moreover, the 150 mM KCl and 75 mM $Na_xH_yPO_4$ conditions resulted in 8 nM and 10 nM EC50's, respectively. The significantly reduced binding in high salt suggests that electrostatic interactions are relatively important to binding and hydrophobic interactions are relatively unimportant.

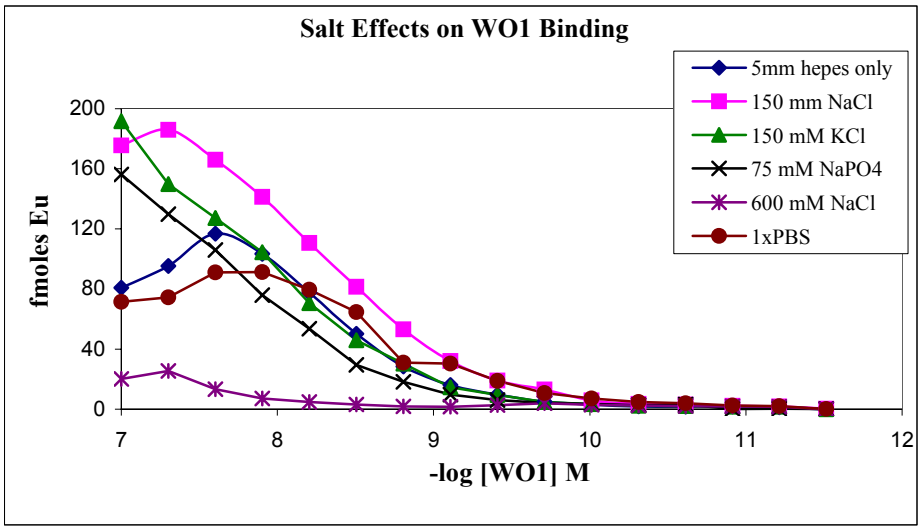



**Figure 7. Salt effect on WO1 binding.** Binding under various salt conditions was calculated from sigmoid midpoints. The $EC_{50}$ values are given in the text.

WO1 binding to Aβ was tested in a range of pH conditions. The Citrate pH 3.0 condition resulted in essentially no binding. Citrate and PIPES at pH 5.8 showed $EC_{50}$ values of 2 nM and >20 nM, respectively. The $EC_{50}$ values for PIPES and Bicine at pH 7.4 were 2 nM and 1.5 nM. For Bicine and CHES at pH 8.5, $EC_{50}$ values were 16 nM and 10 nM. For CHES at pH 10.0, the $EC_{50}$ was > 50 nM. (**Figure 8**). The data suggest that WO1 binds Aβ amyloid best closer to physiological pH values. If hydrophobic interactions dominated WO1- Aβ amyloid binding, this would likely not be the case. This supports the salt effect results suggesting that binding depends at least in part on electrostatic interactions.

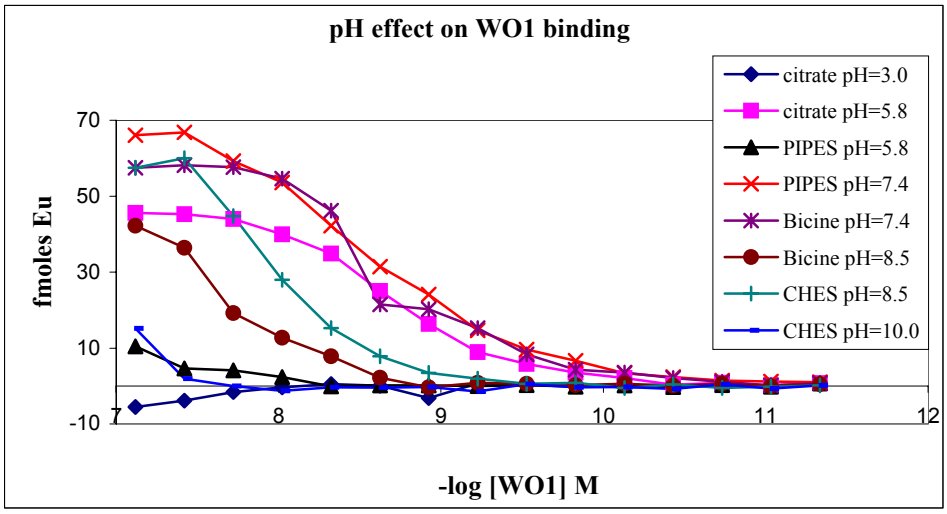



**Figure 8. pH effect on WO1 binding.** The show binding under various pH conditions. Binding affinities were calculated from the sigmoid midpoints. The $EC_{50}$ values are given in the text.

## 5. DISCUSSION

### 5.1 Comparison with Experimental Results

The model of WO1- trigonal prism-model docking that we have proposed, shown in **Figure 5a**, features many hydrogen-bonding pairs and salt-bridges; that agrees with the experimental salt-effect data for WO1- Aβ binding. However, the salt-binding data itself is at odds with WO1's general amyloid-recognition behavior which includes binding to (uncharged) polyglutamine fibrils. This prediction is also called into question by recent data from hydrogen exchange NMR[39] and cysteine scanning mutagenesis[40]. These data are not straightforward to interpret, but they suggest that residues 23-27 (the so-called "B face" of the trigonal-prism protofilament model) may not be in ordered β-sheet structure. There are no strong salt bridges at less than 4.0 Å predicted for binding of WO2 to either the trigonal prism model or the two-strand model.

### 5.2 Assemblies

The Fv models presented here are docked only to a protofilament not to the full Aβ fibril, but these docked models are consistent with the hierarchical fibril model for Aβ amyloid proposed by their authors.[26,38] The fibril model based on the trigonal prism protofilament construction consists of six trigonal prisms packed vertically against one another in two rows of three, in which each protofilament face is exposed twice on the outside surface of the fibril model. For the two-strand model, a full protofilament is generated by applying a two-fold symmetry axis along the C-terminal face perpendicular to the stacking axis of the model. In this way, the open ends of both stacks of monomers point the same way as the stacks twist helically around one another. One depiction of this twisting stack may be found in this work, **Figure 1e**; another is in Petkova, *et al* PNAS 2002, figure 5a. This implies that the C-terminal face of the model is not available for binding; as the predicted docking mode does not employ the C-terminal face, this does not affect our analysis. The Fv model likewise has a single antigen-recognition region for WO1/WO2, whereas the physiological IgM presents ten recognition sites, resulting in high avidity and increased total binding strength.

### 5.3 Implications of Trigonal Prism Results

The docking results for the trigonal prism model suggest that WO1 and WO2 bind to the face of amyloid fibrils similarly. The in-register parallel beta-stacking of the trigonal prism model requires that each residue of the polypeptide line up with its equivalent in the next layer of the prism. Thus, while Glu22 makes peptide bonds to Ala21 and Asp23, it makes hydrogen bonds to Glu22 on neighboring strands. Viewing a face of the prism, then, one would see lines of equivalent residues. The line of Glu22 and Asp23 residues (**Figure 4**) is predicted to bind to the corresponding line of positive and polar residues of WO1's light-chain CDR. The line of Asn27 and Lys28 residues likewise interact with the trail of negative and polar residues.

The lineup of positive and negative charges on the B face with those at the CDR suggests another available binding mode. The A face of the model also has such a lineup of charges (Figure 3). The N-terminal residues of the trigonal prism model occlude part of the A face in this model, which was not allowed torsional freedom, so the A face was less accessible for binding than the B face. Even so, 2/50 solutions for WO1 and 5/50 solutions for WO2, though not clustered, did predict some form of docking to the A face. It is worth noting, furthermore, that earlier simulations with an un-minimized model (in which the N terminal residues took other orientations) yielded a supercluster of docking solutions at the A face of the trigonal prism model with WO1. It is thus likely that a more computationally intensive docking simulation, one which allowed rotation around the bonds of the trigonal prism model, would have shown a bimodal distribution of solutions, one at the A face and another at the B face. These facts, coupled with the new data that suggest non β-sheet conformation for the B-face,[39,40] to some degree weaken the case for the trigonal-prism model for Aβ protofilament. Therefore WO1/WO2 binding to the A face of Aβ remains a distinct possibility. As only 4/50 solutions featured the C face, and they all took very different orientations, our current docking model suggests that the C face is unlikely to feature in WO1 binding.

### 5.4 Implications of Two-Strand Results

Docking calculations based on the two-strand model do not indicate a conclusive result for WO1-binding. However, computations for WO2 support WO2 binding to the end of elongating amyloid fibrils, while only the 48$^{th}$ ranked hit for WO2 and the trigonal prism resembles such an arrangement. This contrasts with expectation that the lines of charge along the N-terminal face would provide a recognition site for the antibodies, but is consistent with the fact that these antibodies bind amyloid fibrils (such as polyglutamine) that lack charged residues. Also, the lower degree

of order found in residues 23-27 in the two-strand model may be more consistent with the new hydrogen exchange and cysteine mutagenesis data.

## 6. CONCLUSIONS

The sequences of two amyloid-recognizing IgMs were compared and analyzed. Structural models of their Fvs were constructed. WO1's Fv was found to be unusually rich in charged residues, while WO2 is rich in Ser residues. As one might expect from the differences in their secondary and ternary structures, the two different protofilament models lead to different predicted docking models. Results of docking simulation for the trigonal-prism model imply that WO1 and WO2 bind to the face of a fibril of Aβ. The results for the two-strand model make no positive prediction for WO1, but results for the two-strand model suggest that WO2 binds to the ends of fibrils. The results for WO1 and the trigonal prism model agree with experimental salt-binding data, while the results for WO2 and the two-strand model are more consistent with WO1/WO2's diverse binding behavior. While these results do not support either model over the other, they make testable predictions.

## 7. ACKNOWLEDGMENTS

We thank Doctors Juntao Guo and Ying Xu for the trigonal prism protofilament model and Doctor Robert Tycko for the two-strand protofilament model.

## 8. REFERENCES

[1] R. H. Falk, R. L. Comenzo and M. Skinner, *N Engl J Med* 337 (1997).
[2] D. J. Selkoe, *Ann N Y Acad Sci* 924 (2000), 17.
[3] I. Tariska, *Acta Med Acad Sci Hung* 21 (1965), 483.
[4] P. A. Temussi, L. Masino and A. Pastore, *Embo J* 22 (2003), 355.
[5] E. T. Jaikaran and A. Clark, *Biochim Biophys Acta* 1537 (2001), 179.
[6] M. R. Hayden and S. C. Tyagi, *Jop* 2 (2001), 124.
[7] O. Corti and A. Brice, *J Soc Biol* 196 (2002), 95.
[8] R. Kruger, O. Eberhardt, O. Riess and J. B. Schulz, *Trends Mol Med* 8 (2002), 236.
[9] J. M. George, *Genome Biol* 3 (2002), REVIEWS3002.
[10] M. Pras, F. Prelli, E. C. Franklin and B. Frangione, *Proc Natl Acad Sci U S A* 80 (1983), 539.
[11] H. D. Feiner, *Hum Pathol* 19 (1988), 1255.
[12] J. N. Buxbaum, J. V. Chuba, G. C. Hellman, A. Solomon and G. R. Gallo, *Ann Intern Med* 112 (1990), 455.
[13] V. Bellotti, P. Mangione and G. Merlini, *J Struct Biol* 130 (2000), 280.
[14] C. A. Ross, *Neuron* 35 (2002), 819.
[15] C. Weissmann, *Nat Rev Microbiol* 2 (2004), 861.
[16] H. Inouye, P. E. Fraser and D. A. Kirschner, *Biophys J* 64 (1993), 502.
[17] M. Sunde and C. Blake, *Adv Protein Chem* 50 (1997), 123.
[18] S. B. Malinchik, H. Inouye, K. E. Szumowski and D. A. Kirschner, *Biophys J* 74 (1998), 537.
[19] A. D. Williams, et al., *J Mol Biol* 335 (2004), 833.
[20] J. D. Harper, S. S. Wong, C. M. Lieber and P. T. Lansbury, Jr., *Biochemistry* 38 (1999), 8972.
[21] L. C. Serpell, et al., *J Mol Biol* 300 (2000), 1033.
[22] T. C. Saido, W. Yamao-Harigaya, T. Iwatsubo and S. Kawashima, *Neurosci Lett* 215 (1996), 173.
[23] I. Kheterpal, A. Williams, C. Murphy, B. Bledsoe and R. Wetzel, *Biochemistry* 40 (2001), 11757.
[24] O. N. Antzutkin, R. D. Leapman, J. J. Balbach and R. Tycko, *Biochemistry* 41 (2002), 15436.
[25] M. Torok, et al., *J Biol Chem* 277 (2002), 40810.
[26] A. T. Petkova, et al., *Proc Natl Acad Sci U S A* 99 (2002), 16742.
[27] T. S. Burkoth, et al., *J. Am. Chem. Soc.* 122 (2000), 7883.
[28] I. Kheterpal, et al., *Biochemistry* 42 (2003), 14092.
[29] S. S. Wang, S. A. Tobler, T. A. Good and E. J. Fernandez, *Biochemistry* 42 (2003), 9507.
[30] M. Kraus, M. Bienert and E. Krause, *Rapid Commun Mass Spectrom* 17 (2003), 222.
[31] I. Kheterpal, S. Zhou, K. D. Cook and R. Wetzel, *Proc Natl Acad Sci U S A* 97 (2000), 13597.
[32] T. L. Benzinger, et al., *Proc Natl Acad Sci U S A* 95 (1998), 13407.
[33] M. F. Perutz, J. T. Finch, J. Berriman and A. Lesk, *Proc Natl Acad Sci U S A* 99 (2002), 5591.
[34] L. O. Tjernberg, et al., *Biochem J* 366 (2002), 343.

[35] C. Hilbich, B. Kisters-Woike, J. Reed, C. L. Masters and K. Beyreuther, *J Mol Biol* 218 (1991), 149.
[36] M. O. Chaney, S. D. Webster, Y. M. Kuo and A. E. Roher, *Protein Eng* 11 (1998), 761.
[37] L. C. Serpell, C. C. Blake and P. E. Fraser, *Biochemistry* 39 (2000), 13269.
[38] J. T. Guo, R. Wetzel and Y. Xu, *Proteins* 57 (2004), 357.
[39] N. A. Whittemore, et al., *Biochemistry* 44 (2005), 4434.
[40] S. Shivaprasad and R. Wetzel, *Biochemistry* 43 (2004), 15310.
[41] R. Nelson, et al., *Nature* 435 (2005), 773.
[42] B. O'Nuallain and R. Wetzel, *Proc Natl Acad Sci U S A* 99 (2002), 1485.
[43] M. Dumoulin and C. M. Dobson, *Biochimie* 86 (2004), 589.
[44] A. C. Davis and M. J. Shulman, *Immunol Today* 10 (1989), 118.
[45] N. R. Whitelegg and A. R. Rees, *Protein Eng* 13 (2000), 819.
[46] Chemical Computing Group Inc., 1997.
[47] E. E. Abola, J. L. Sussman, J. Prilusky and N. O. Manning, *Methods Enzymol* 277 (1997), 556.
[48] J. L. Sussman, et al., *Acta Crystallogr D Biol Crystallogr* 54 (1998), 1078.
[49] A. C. Martin, *Proteins* 25 (1996), 130.
[50] S. F. Altschul, et al., *Nucleic Acids Res* 25 (1997), 3389.
[51] C. Chothia and A. M. Lesk, *J Mol Biol* 196 (1987), 901.
[52] T. T. Wu and E. A. Kabat, *Mol Immunol* 29 (1992), 1141.
[53] A. C. Martin, J. C. Cheetham and A. R. Rees, *Proc Natl Acad Sci U S A* 86 (1989), 9268.
[54] A. C. Martin, J. C. Cheetham and A. R. Rees, *Methods Enzymol* 203 (1991), 121.
[55] A. R. Rees, Searle,S.M.J., Henry,A.H.and Pedersen,J.T., *in* "Protein Structure Prediction" (S. M.J.E., ed.), p. 141–172. Oxford University Press, Oxford, 1996.
[56] J. T. Pedersen, Searle,S.M.J., Henry,A.H. and Rees,A.R., *Immunomethods* 1 (1992), 126.
[57] G. M. Morris, et al., *Journal of Computational Chemistry* 19 (1998), 1639.
[58] S. J. K. Weiner, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P., *J Am Chem Soc* 106 (1984), 765.
[59] S. J. K. Weiner, P. A.; Nguyen, D. T.; Case, D. A., *J Comput Chem* 7 (1986), 230
[60] Accelrys; MSI, San Diego, 2000.
[61] G. Murshudov, *Acta Crystallogr D Biol Crystallogr.* 53 (1997), 240.
[62] N. Collaborative Computational Project, *Acta Crystallogr D* 50 (1994), 760.
[63] C. Cole and J. Warwicker, *Protein Sci* 11 (2002), 2860.
[64] W. L. DeLano. DeLano Scientific, San Carlos, CA, USA, 2002.
[65] A. Szilagyi and P. Zavodszky, *Structure Fold Des* 8 (2000), 493.
[66] Y. Li, H. Li, S. J. Smith-Gill and R. A. Mariuzza, *Biochemistry* 39 (2000), 6296.
[67] C. G. Dealwis, et al., *J Mol Biol* 236 (1994), 342.
[68] V. M. Reyes, M. R. Sawaya, K. A. Brown and J. Kraut, *Biochemistry* 34 (1995), 2710.

# SNAPSHOTS OF THE PANTOTHENATE SYNTHETASE FROM MYCOBACTERIUM TUBERCULOSIS ALONG THE REACTION COORDINATE

Shuishu Wang
*Public Health Research Institute, 225 Warren Street, Newark, New Jersey 07103*

David Eisenberg
*Howard Hughes Medical Institute, UCLA-DOE Institute of Genomics and Proteomics, Molecular Biology Institute, UCLA, Box 951570, Los Angeles, CA 90095-1570*

## 1. ABSTRACT

Pantothenate synthetase (PS) from *Mycobacterium tuberculosis* represents a potential target for developing new anti-tuberculosis drugs. PS catalyzes the ATP-dependent condensation of pantoate and β-alanine to form pantothenate. We determined the crystal structures of PS from *M. tuberculosis* and its complexes with substrates ATP, pantoate, and β-alanine, as well as an ATP analog AMPCPP, a reaction intermediate pantoyl adenylate and a reaction product AMP, with resolutions from 1.6 to 2 Å. The PS structure reveals a dimer, and each subunit has two domains with tight association between domains. The active site cavity is on the N-terminal domain, covered by the C-terminal domain. There is a flexible loop that forms one wall of the active site cavity, and it opens and closes the active site cavity. The enzyme binds ATP and pantoate tightly in the active site, and brings the carboxyl oxygen of pantoate near the α-phosphorus atom of ATP for an in-line nucleophilic attack. The pantoyl adenylate thus formed is stabilized by tight binding interactions with enzyme active site residues, and protected from hydrolysis by closing of the active site cavity with the flexible loop becoming ordered. Binding of β-alanine can occur only after formation of the pantoyl adenylate intermediate. Crystal structures of these complexes provide a step-by-step view of the PS catalyzed reaction, and allow us to construct transient reaction intermediates and deduce detailed reaction mechanism. PS catalyzes the reaction by stabilizing the reaction intermediates along the reaction coordinate, and structural models of these intermediates can provide basis for inhibitor design.

## 2. INTRODUCTION

Pantothenate (vitamin B5) is an essential precursor for the biosynthesis of coenzyme A and acyl carrier proteins, both of which play critical roles in many cellular processes including energy metabolism and fatty acid metabolism [1]. Microorganisms and plants can synthesize pantothenate, while animals obtain this essential nutrient from their diet [2]. Therefore, the pantothenate biosynthetic pathway offers targets for developing drugs against microbial pathogens. Recently, Jacobs and coauthors [3] reported that a *Mycobacterium tuberculosis* (MTB) mutant defective in the de novo biosynthesis of pantothenate is highly attenuated in both immunocompromised and immunocompetent mice. This observation indicates that a functional pantothenate biosynthetic pathway is essential for virulence of MTB, and thus represents a new target for anti-tuberculosis agents.

The pantothenate biosynthetic pathway in bacteria comprises four steps catalyzed by enzymes encoded by the *panB*, *panC*, *panD* and *panE* genes [4]. The *panC* gene encodes a pantothenate synthetase (PS), which catalyzes the last step of pantothenate biosynthesis, the ATP-dependent condensation of pantoate and β-alanine to form pantothenate. The *panC* gene product in several organisms has been identified and characterized, including those in *Escherichia coli* [5], *Saccharomyces cerevisiae*, higher plants *Oryza sativa* and *Lotus japonicus* [6], fungus *Fusarium oxysporum* [7], and *M. tuberculosis* [8]. The PS enzymes from *E. coli*, higher plants, and MTB form dimers in solution.

The *E. coli* PS enzyme structure belongs to the cytidylyltransferase superfamily [5]. It has two distinct domains, a large N-terminal domain having a Rossmann fold and a smaller C-terminal domain containing a helical layer above a three-stranded antiparallel β-sheet. Based on structural comparison of the *E. coli* PS with other members of cytidylyltransferase superfamily having known structures, von Delft et al. [5] deduced the ATP and pantoate binding sites of the *E. coli* PS and proposed a hinged domain mechanism for opening and closing of the enzyme active site cavity.

Kinetic analysis of the MTB PS suggests that the enzyme-catalyzed reaction proceeds through two steps: the formation of an enzyme bound intermediate, pantoyl adenylate, from ATP and pantoate, followed by nucleophilic attack on the intermediate by β-alanine to form pantothenate and AMP [8]. The existence of pantoyl adenylate as an enzyme bound intermediate is suggested by the transfer of the $^{18}O$ label from the carboxyl group of pantoate to the reaction product AMP. However, pantoyl adenylate cannot be directly isolated or identified by TLC methods, presumably due to rapid lactonization [9].

Here, we present the crystal structure of the MTB PS enzyme and structures of its complexes with substrates, a substrate analog (AMPCPP), a reaction product (AMP), and a pantoyl adenylate reaction intermediate [10, 11]. The MTB PS has the same fold as the *E. coli* enzyme. However, in the MTB PS structure the domains of each subunit have a closed conformation, in contrast to those of the *E. coli* PS structure [5], and there is no significant movement between domains among all structures of the apo-enzyme and various complexes. A flexible region, which forms a wall of the active site cavity, becomes ordered in the reaction intermediate complex and closes the active site cavity, thus acting as a gate to the active site cavity of the MTB PS enzyme. The structure of the β-alanine complex gives direct evidence that the binding site for this last substrate exists only after formation of the pantoyl adenylate intermediate. Together these crystal structures lead to a detailed view of the overall reaction mechanism. Models of the transient reaction intermediate can be easily constructed from the experimentally observed complexes.

## 3. MATERIALS AND METHODS

### 3.1 Protein Production

The detailed procedures for cloning of the *panC* gene, expression and purification of the protein have been described [10]. Briefly, the MTB *panC* gene (Rv3602c) encoding the pantothenate synthetase was amplified from the genomic DNA of MTB strain H37Rv and inserted into a pET30a plasmid (Novagen), which produces a recombinant protein with an N-terminal 6xHis tag that can be cleaved off with enterokinase. BL21(DE3) cells containing the pET30-*panC* plasmid were grown in LB medium containing 50 μg/ml kanamycin at 37 °C and induced with 0.4 mM IPTG for 3 hours. Cells were collected and lysed by lysozyme treatment followed by sonication in 20 mM HEPES pH 7.8, 500 mM NaCl, and 0.5 mM PMSF. The protein was purified from a $Ni^{2+}$-charged HiTrap chelating column (Amersham) with 20 mM HEPES pH 7.8, 500 mM NaCl and a linear gradient of imidazole. The PS protein was subjected to enterokinase (New England Biolabs) digestion to cleave off the N-terminal fusion tag. Electrospray mass spectroscopy indicated that 9 residues from the C-terminus of the protein were also cleaved off by enterokinase digestion. However, enzyme activity assays suggested that this C-terminal truncation does not have any detectable effect on the enzyme activity [10]. The enterokinase digested protein was further purified with a second round of chromatography on a $Ni^{2+}$-HiTrap column followed by Superdex 75 column (Amersham) gel filtration.

### 3.2 Crystallization and Data Collection

Crystallization was carried out as described previously [10]. Crystals in space group $P2_1$ were obtained from drops set up with well solutions containing 10-15% PEG 3000, 5% glycerol, 2% ethanol, 20 mM $MgCl_2$, 150 mM $Li_2SO_4$ and 100 mM imidazole pH 8.0 at 20 ºC. Those in space group C2 were from similar conditions, except with 2% isopropanol in place of ethanol and with 200 mM $Li_2SO_4$. Cocrystallization with β-alanine was done with a well solution of 15% PEG 3000, 2% ethanol, 20 mM $MgCl_2$, 100 mM imidazole pH 8.0 and 20 mM β-alanine, from which crystals in space group $P2_12_12_1$ were obtained. However, these crystals do not have β-alanine in the active site. Cocrystallization with both ATP and pantoate was carried out in the same crystallization condition of the $P2_1$ crystals but with both ATP and pantoate added, from which isomorphous $P2_1$ crystals were obtained that have a pantoyl adenylate intermediate in the active site.

Crystals in complex with substrates were obtained by soaking apo-enzyme crystals in solutions containing substrates or the ATP analog AMPCPP. Soaking experiments were carried out by adding solutions containing substrates or AMPCPP directly to the drops, or to drops of pseudo mother liquor with transferred crystals, and incubating overnight or longer. Crystals were soaked in solutions containing ~10 mM each of AMPCPP alone, both AMPCPP and pantoate, both pantoate and β-alanine, both ATP and pantoate, and both AMP and β-alanine. Crystals were also soaked in solutions containing up to 20 mM pantothenate, but no pantothenate was found in the crystal structure.

Before data collection, crystals were soaked for 2-5 min in a cryogenic solution similar to well solutions or substrate soaking solutions with glycerol added to 30%, and they were then flash-frozen in a cryo stream of $N_2$ gas at 100 K. Diffraction data were collected at 100 K on a Rigaku FRD generator with an R-AXIS $IV^{++}$ detector. Data reduction and scaling were carried out with the programs DENZO and SCALEPACK [12]. Data processing statistics are reported in references [10] and [11]. Table 1 summarizes a few of the data sets, of which the structures are mentioned in this paper. The $P2_1$ crystal has two PS molecules per asymmetric unit; while the C2 crystal contains one molecule per asymmetric unit. In the C2 crystal, the active site is locked in an inactive conformation by the crystal packing (see below), and the pantoate and β-alanine binding sites are compromised. Therefore, only AMP was found when we soaked the C2 crystals with both AMP and β-alanine (data set 8, in Table 1), and those soaked with both ATP and pantoate gave an ATP complex (data set 9) instead of a reaction intermediate complex.

**TABLE 1**

Summary of crystal structure data and experiments

| | Complex | Experiment | SG | #/A.U. | resolution | PDB ID | Ref. |
|---|---|---|---|---|---|---|---|
| 1 | Apo enzyme | | $P2_1$ | 2 | 1.6 Å | 1MOP | [10] |
| 2 | AMPCPP and pantoate in A; AMPCPP in B | AMPCPP, pantoate, soak | $P2_1$ | 2 | 1.6 Å | 1N2E | [10] |
| 3 | Pantoyl adenylate | ATP, pantoate, cocrystal | $P2_1$ | 2 | 2.0 Å | 1N2H | [10] |
| 4 | Pantoyl adenylate | ATP, pantoate, soak | $P2_1$ | 2 | 1.7 Å | 1N2I | [10] |
| 5 | Pantoate | Pantoate, β-alanine, soak | $P2_1$ | 2 | 1.8 Å | 1N2J | [10] |
| 6 | AMP, β-alanine in A; AMP only in B | AMP, β-alanine soak | $P2_1$ | 2 | 1.85 Å | 1A86 | [11] |
| 7 | Apo enzyme | | C2 | 1 | 1.7 Å | 2A88 | [11] |
| 8 | AMP | AMP, β-alanine soak | C2 | 1 | 1.7 Å | 2A7X | [11] |
| 9 | ATP | ATP, pantoate, soak | C2 | 1 | 1.6 Å | 2A84 | [11] |

### 3.3 Structural Determination and Refinement

The crystal structure of the apo-enzyme in the $P2_1$ crystal was determined [10] by the molecular replacement method using AMORE [13], with subunit A of the *E. coli* PS structure (PDB ID code 1IHO [5]) as a model. The two domains were separated as two independent search models. The positions of the two N-terminal domains and one C-terminal domain were found from the rotation and translation searches, and the other C-terminal domain was generated by a non-crystallographic-symmetry (NCS) operation on the position of the known C-terminal domain with the NCS symmetry operators derived from the two N-terminal domains. The structure of the apo-enzyme was used to calculate phases for data sets of various complexes by rigid body refinement into new data sets for isomorphous crystals, or by molecular replacement with AMORE. The final refinement statistics are reported in references [10, 11]. Structural models were refined against diffraction data using CNS [14] and/or REFMAC [15]. After each cycle of refinement, models were manually adjusted with electron density maps using O [16]. The same subsets of data for $R_{free}$ calculation were kept between the CNS and REFMAC programs.

Besides the 9 residues that were digested off by enterokinase, there are 10 more residues from the C-terminus that do not have clear electron density. Based on its susceptibility to protease digestion and lack of clear electron density, it is reasonable to assume that these residues at the C-terminus of the protein are flexible, and do not have a defined structure in solution. It is clear from the structure that both N- and C-termini are away from the active site cavity, and therefore are unlikely to affect the enzyme catalytic activity. This is consistent with the results of enzyme activity assays [10].

## 4. STRUCTURE OF THE MTB PANTOTHENATE SYNTHETASE

### 4.1 Dimer Structure

The crystal structure of the MTB PS enzyme indicates that it is a dimer (**Figure 1**), resembling a butterfly when viewed from one side. The dimer interface is extensive, having a buried area of about 2150 $Å^2$. A cluster of 8 hydrophobic side chains at the center of the dimer interface, Phe174, Leu177, Leu144, and Val118, form the core of the dimer interaction. This cluster is surrounded by hydrogen bonds and salt bridges, as well as the specific β strand

interactions of the two-stranded intersubunit β sheet that is composed of the strand β5 from each subunit. The MTB PS enzyme also exists in solution as a dimer [8, 10], as do PS enzymes from higher plants [6, 7].

The two subunits of the dimer in the $P2_1$ crystal are similar to each other, with an rmsd of 0.9 for all Cα atoms except the disordered residues. Most of the deviation lies at residues 74 to 88, and residues 260 to 265, where the crystal packing environments are different. Residues 74 to 88 (β3-$3_{10}$3-α3'-α3) form one wall of the active site cavity (see **Figure 2** and the active site cavity section b 31 In subunit A these residues are ordered with residues 74 to 76 having high B factors, while in subunit B re- 75 to 80 (helix α3'and nearby residues) do not have electron density, and the rest of the residues up to residue 88 are partially disordered with high B factors. This flexible wall of the active site cavity becomes ordered in the complex with the pantoyl adenylate reaction intermediate, suggesting their function as a gate to the active site cavity (see below pantoyl adenylate binding interactions).

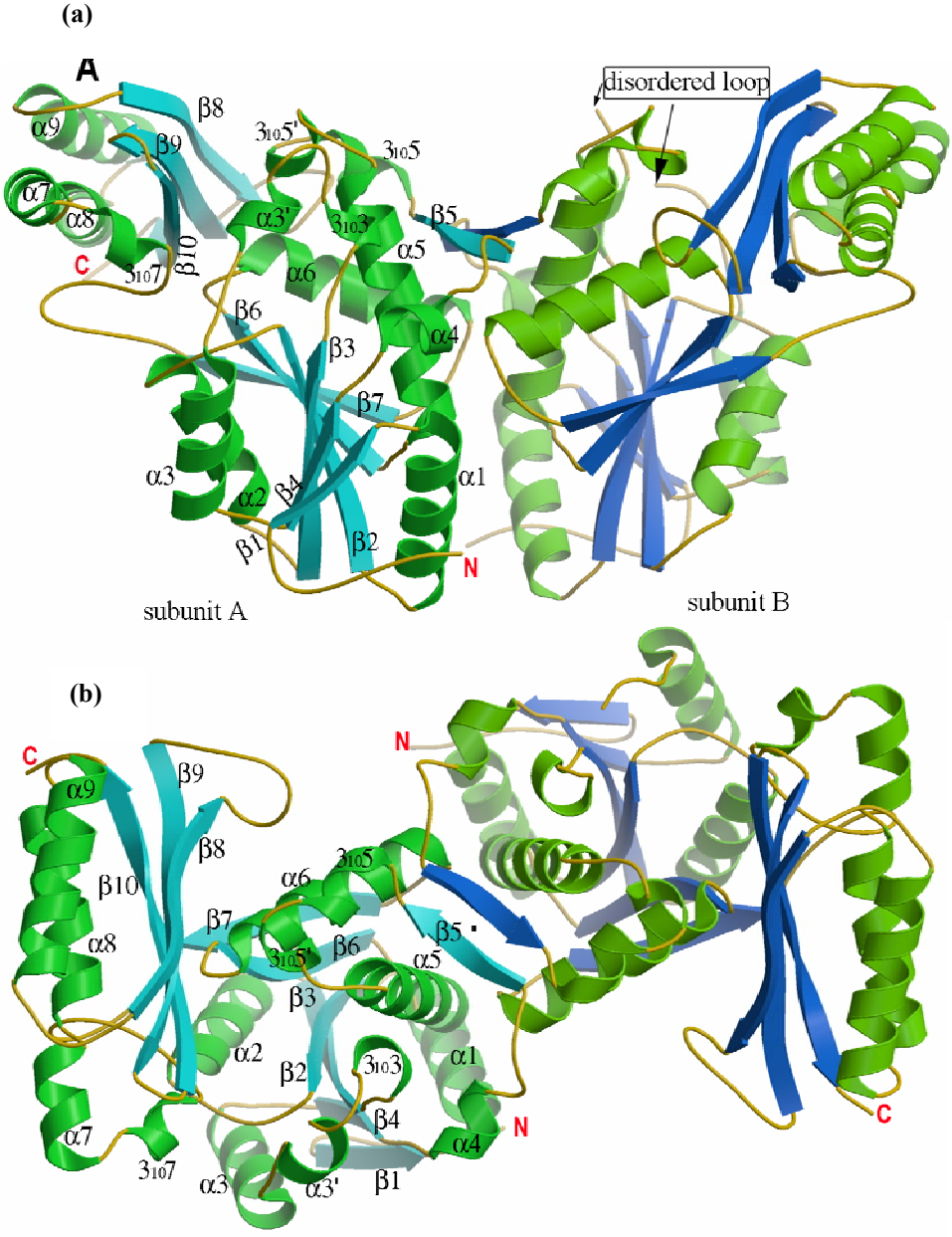



**Figure 1.** Ribbon diagram of the *M. tuberculosis* pantothenate synthetase dimer. **(a)** A side view of the dimer structure shows that it resembles the shape of a butterfly. **(b)** An orthogonal view of (A) from top, with the two-fold NCS symmetry axis (labeled with a dot) approximately perpendicular to the paper plane. Secondary structure elements for subunit A (left) are labeled. Those for subunit B are identical except that the short helix α3' is disordered. The figure was prepared from the apo-enzyme structure of the $P2_1$ crystal, with Molscript [17] and Raster3D [18].

The C2 crystal differs from the $P2_1$ crystal in that there is one polypeptide chain per asymmetric unit [11]. However, the protein molecule in the C2 crystal is also present as a dimer that is essentially identical to that in the $P2_1$ crystal, but the two-fold axis of the dimer coincides with one crystal symmetry axis. In the C2 crystal, a larger portion of the flexible wall, from residue 76 to 86, is completely disordered. The first turn of helix α3 unwinds, and helix $3_{10}3$, which contains residue Glu72 that is important for binding pantoate (see below pantoate binding site), also unwinds and forms an extended structure. This 32 ral difference is induced by crystal packing and the intrinsic flexibility of the loop.

### 4.2 Domain Interactions

Each subunit of the dimer has two well-defined domains (**Figure 1**). The N-terminal domain has a Rossmann fold, with a central parallel β sheet composed of strands β1 to β4, β6, and β7, and helices on both sides of the β-sheet. Strand β7 leads to the C-terminal domain, which starts with a hairpin loop followed by helix $3_{10}7$. The rest of the C-terminal domain has a simple two-layer structure: a layer of helices α7, α8, and α9 followed by a layer of antiparallel β sheet composed of strands β8 to β10.

The two domains have extensive interactions. There is a hydrophobic core composed of residues Leu123, Leu127, Tyr162, Val166 in the N-terminal domain and residues Tyr249, Leu257, and Leu269 in the C-terminal domain. Surrounding this hydrophobic core are three salt bridges, Glu159 to Arg267, Glu128 to Arg278, Glu126 to Arg253, and 6 additional hydrogen bonds linking the two domains. In addition, there are several water-mediated hydrogen bonding interactions. The relative positions of the two domains stay the same for the two subunits in the dimer and for all crystals including the $P2_1$, $P2_12_12_1$, and C2 crystals. These structural features suggest that domain hinge motion is unlikely to be the mechanism of opening and closing the active site cavity for MTB PS.

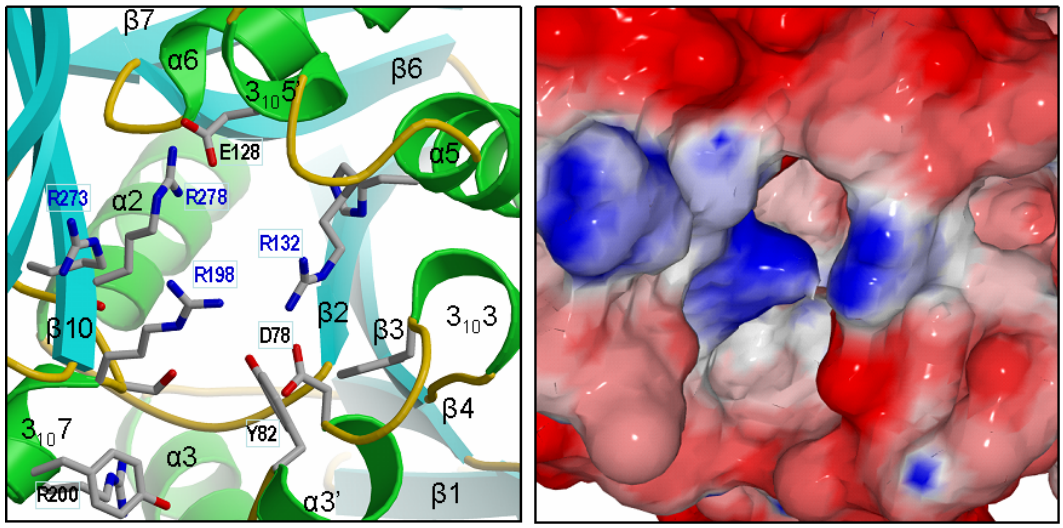



**Figure 2.** Active site cavity and the electrostatic potential surface around the entrance to the active site cavity. The figure was prepared from the subunit A of the $P2_1$ crystal. The two panels are approximately the same view. In the left panel, side chains at the entrance to the activity are shown as sticks. The segment from the end of β3 to the beginning of α3 forms the flexible wall, which is disordered in subunit B and in the C2 crystal thus opening the active site cavity.

### 4.3 Active Site Cavity

Typical for nucleotide binding (Rossmann fold) proteins, the active site is located at the C-terminus of the central parallel β sheet. The cavity is large and deep, and its bottom is at a cleft between strands β2 and β6 (**Figure 2**). The segment between β3 and α3 (loop-$3_{10}$3-loop-α3'-loop region) forms one side of the walls surrounding the cavity in subunit A of the $P2_1$ crystal. Part of this loop is disordered in subunit B, thus opens up the active site cavity. Other walls of the cavity are from β2-loop-α2, $3_{10}5'$-loop-α5, β6-loop-α6, and the loop after β7. Helix $3_{10}7$ and the β strands from the C-terminal domain partially cover the top of the active site cavity. The bottom of the active site cavity is mainly hydrophobic, while the top half of the cavity has several charged residues, including His44 and His47 at the N-terminus of α2 (the HIGH motif [5, 19]), Lys160 and Asp161 at the N-terminus of α6, and Arg198 of $3_{10}7$ (KMSKS motif [5]). At the entrance to the active site cavity, several charged side chains stick to the center,

leaving only a small opening to the cavity (**Figure 2**, right panel). Four arginine side chains, including Arg198, form a patch of positive charges at the entrance. These positive charges may steer the negatively charged substrates into the active site cavity. As we will see later, Arg198 plays an important role in binding the phosphate groups of ATP.

### 4.4 ATP Binding Interactions

The active site is well designed for binding ATP. **Figure 3** shows the initial difference electron density for the ATP molecule in the active site of the C2 crystal. The ATP molecule has an average temperature factor of 17.3 $Å^2$, virtually identical to the average temperature factor of the surrounding protein atoms, indicating that it is at full occupancy. A magnesium ion was found binding to the phosphate groups of ATP. It has a nearly perfect octahedral coordination. Its distances to the O2α, O2β, and O1γ atoms of the phosphate groups are 2.26 Å, 1.99 Å, and 2.12 Å, respectively; three other ligands are water molecules with distances of 1.96 Å, 2.11 Å, and 2.28 Å. The adenosine group of ATP binds tightly at the bottom of the active site cavity with hydrogen bonds and hydrophobic interactions [10, 11]. The phosphate groups stick toward the top of the active site cavity, and are located near the N-terminal ends of helices α2 and $3_{10}7$. Binding of the phosphate groups involves mainly main chain NH groups and charged side chains.

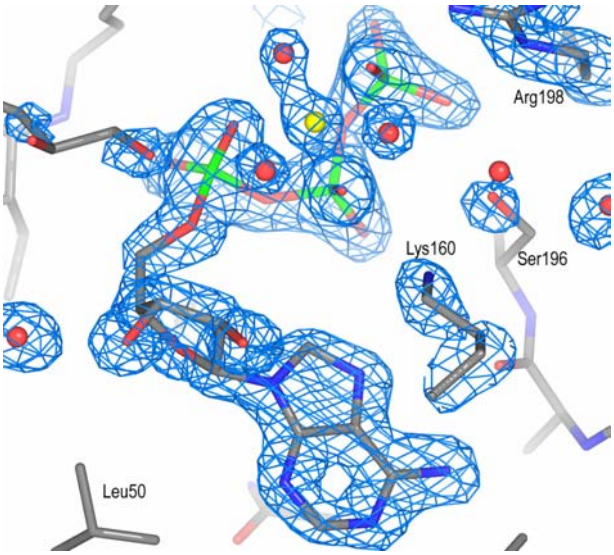



**Figure 3**. Initial Fo – Fc electron density in the active site of the C2 crystal soaked in a solution containing ATP and pantoate superimposed on the final refined model. The map was calculated at 1.6 Å and contoured at 3.5 σ. The ATP molecule has very well defined electron density. Side chains of Lys160, Ser196 and Arg198 moved relative to those in the apo-enzyme to interact with the phosphate groups, and thus have positive initial difference electron density. The figure is prepared with PYMOL [20].

Two His residues of the HIGH motif [5, 19, 21] interact with the phosphates of ATP. The first histidine, His44, has a hydrogen bond extending from its Nε2 to O1β of β-phosphate. The His47 side chain has a hydrogen bond from its Nε2 to the bridge oxygen between the α- and β-phosphate, suggesting that it functions as a general acid by donating a proton to the leaving pyrophosphate group. Both histidine residues have the Nδ1 atom forming a hydrogen bond to a main chain NH group, and thus the Nε2 atoms are hydrogen bond donors. Their imidazole rings are in close proximity to each other with the closest separation being ~3.2 Å, indicating favorable π-π interactions. These two side chains also play a role in binding the adenine group and stabilizing the reaction intermediate, because the imidazole rings are close to the adenine (closest distances of ~3.4 and ~4.3 Å for His44 and His47, respectively). Mutation of either histidine residue to alanine causes a greater than 1000-fold reduction in enzyme activity [22]. The Gly46 residue is highly conserved in this motif. It flanks one side of adenine (with Lys160 on the loop after β6 on the other side), where any side chain larger than a hydrogen atom would have steric hindrance. The KSMKS motif [23] is not conserved at the sequence level in PS enzymes, but it is evident from structural alignments [5]. Residues Ser196, Ser197, and Arg198, which are at the N-terminus of helix $3_{10}7$ and are part of the KSMKS motif, bind the β- and γ-phosphate groups. The side chains of Lys160 and Arg198 are partially disordered in the apo-enzyme structure, but become ordered and move to enable salt-bridges to the β- and γ-phosphate groups, respectively. The side chain of Lys160 also packs against one side of adenine. Mutation of this

residue to alanine reduces the enzyme activity more than 1000-fold, and it decreases affinity of the enzyme for ATP [22]. Another residue that binds ATP is Asp161, which has hydrogen bonds to two water ligands of the magnesium ion, and to O2* of the ribose group. Overall, the ATP molecule is rigidly held in the active site by many hydrophobic, π-electron, hydrogen bonding, and charge-charge interactions, which keep it in an optimal conformation for a nucleophilic attack by pantoate. Residues involved in binding the adenosine group are also important in stabilizing the pantoyl adenylate intermediate.

The binding interactions of ATP with protein are similar to those of AMPCPP [10]. However, the methylene group in AMPCPP replaces the bridge oxygen between the α- and β-phosphate groups of ATP, and thus it is not capable of forming a hydrogen bond with the Nε2 atom of His47. Consequently, the α-phosphate moves slightly away from amide nitrogen of Met40, and thus no hydrogen bond is formed between them in the AMPCPP complex. Therefore, ATP has slightly higher binding affinity to the enzyme active site than its non-reactive analog AMPCPP.

**4.5 Pantoate Binding Site**

In the $P2_1$ crystal that was soaked with pantoate and β-alanine, there is a full occupancy of pantoate in both active sites [10]. The binding site for pantoate is located in a pocket at the bottom of the active site cavity, next to the binding site of the adenosine group (**Figure 4a**). This is the same binding site proposed for the *E. coli* PS protein [5], although the pantoate molecule has a different conformation and therefore different binding interactions. The pantoate molecule is tightly bound at the bottom of the active site cavity, with side chains of two glutamine residues, Gln72 from helix $3_{10}3$ and Gln164 from helix α3, forming hydrogen bonds to its hydroxyl groups and one carboxyl oxygen. The two methyl groups face the hydrophobic groups at the bottom of the active site cavity, Pro38 and Phe157 side chains. The side chain of Met40 packs against the hydrophobic side of the pantoate molecule. Mutation of either glutamine residue to alanine greatly reduces the rate of pantoyl adenylate formation [22].

Soaking $P2_1$ crystals in solutions containing both AMPCPP and pantoate gave a full occupancy of AMPCPP, but with a glycerol in the pantoate binding site in subunit B, which has the flexible wall (residues 75-88) disordered. However, subunit A has a partial occupancy of both AMPCPP and pantoate, as indicated by their electron density and B factors. Increasing the concentration of pantoate in the soaking solution gave only a slightly higher occupancy of pantoate but a lower occupancy of AMPCPP [10]. All these observations indicate that AMPCPP and pantoate cannot coexist in the active site. This must be due to the steric and charge repulsions between the α-phosphate group and the carboxyl group of pantoate. Modeling both ATP and pantoate in one active site based on the fully occupied complexes will put one carboxyl oxygen atom of pantoate ~2.7 Å from the α-phosphorus atom of ATP [11]. This is essentially identical to what is observed in subunit A that has partial occupancy of both AMPCPP and pantoate. The carboxyl oxygen of pantoate is in a good position for an in-line nucleophilic attack on the α-phosphate. Because residues of the flexible wall (gate residues) are disordered in subunit B, the active site cavity is open and allows AMPCPP to diffuse easily into the active site. The extensive binding interactions of AMPCPP to the protein atoms make its binding thermodynamically more favorable than the binding of pantoate to the protein. Therefore, AMPCPP has a full occupancy in the active site cavity of subunit B of the complex with both AMPCPP and pantoate. On the other hand, the gate residues are less flexible in the subunit A due to crystal packing, thus making it kinetically less favorable for AMPCPP to bind, and allowing pantoate to have a partial occupancy in the active site of subunit A.

In the C2 crystal, the side chain of Gln72 on helix $3_{10}3$ swings out of the active site due to crystal packing, and the binding of pantoate is abolished. We observed only ATP in the active site with a full occupancy (see above ATP binding interactions) when we soaked the C2 crystal in a solution containing both ATP and pantoate [11]. There was no electron density for pantoate or pantoyl adenylate in the active site cavity. In the pantoate binding pocket, there was positive difference electron density, into which we modeled a glycerol molecule. This glycerol has a high temperature factor, suggesting that it has a low occupancy. Except for this Gln72 side chain, the rest of the binding site for pantoate is intact. In the $P2_1$ crystal, soaking with both ATP and pantoate results in a pantoyl adenylate intermediate in the active site cavity, and the disordered loop in subunit B becomes ordered [10] (see more details below). This indicates that loss of this glutamine side chain completely abolishes the enzyme catalytic activity for the first half of the reaction inside the crystal. The glutamine side chains are also involved in binding the reaction intermediate. Thus mutation of either residue could affect the stability of pantoyl adenylate intermediate. It is noteworthy that Asn69 was also found to be important for the enzyme activity [22]. This residue is not directly involved in binding substrates or the intermediate. However, it forms a hydrogen bond to the side chain of Gln72, and thus is important for keeping the Gln72 side chain in position. In addition, since both Asn69 and Gln72 are on

the one-turn $3_{10}3$ helix, mutation of Asn69 might also affect the stability of this $3_{10}$ helix and hence the position of the Gln72 side chain.

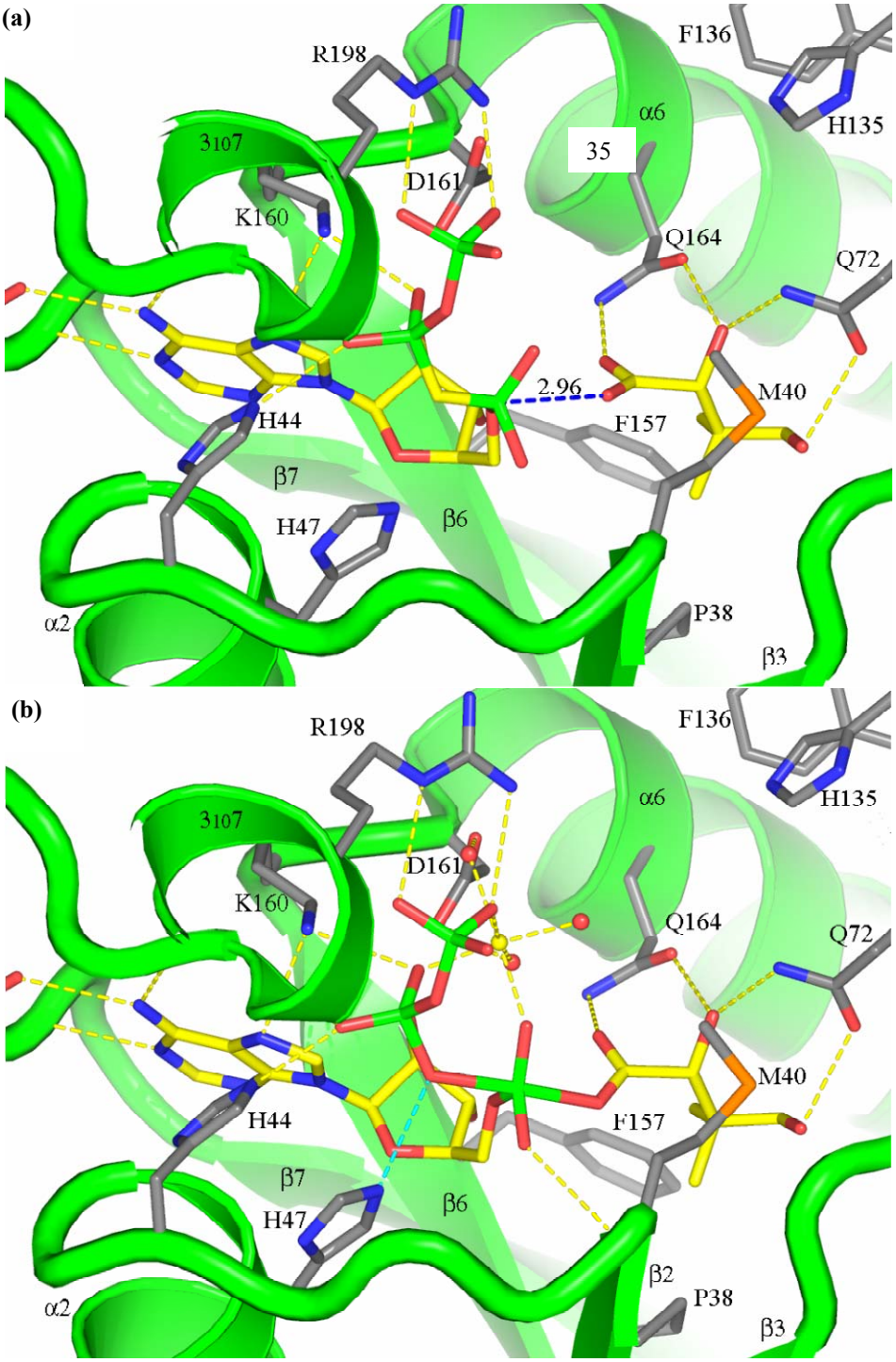



**Figure 4**. **(a)** Binding interactions of AMPCPP and pantoate in the active site of subunit A of the $P2_1$ crystal soaked in a solution containing both AMPCPP and pantoate. Both molecules have a partial occupancy (see text for details), and the magnesium ion is not well defined in this active site. However, binding position and interactions are essentially identical to those of AMPCPP and pantoate with a full occupancy when each is bound in the active site alone. The distance between the α-phosphorus atom and the carboxyl oxygen is ~2.96 Å, shown in blue dashed line. **(b)**

A model of the trigonal bipyramidal intermediate based on crystal structures of the ATP complex, AMPCPP complex and pantoate complex. Formation of the intermediate requires only the movement of the α-phosphorus atom towards the carboxyl oxygen. All favorable interactions with active residues are preserved. His47 is poised to function as a general acid.

### 4.6 Binding Interactions of the Reaction Intermediate, Pantoyl Adenylate

Soaking the $P2_1$ crystals with solutions containing both ATP and pantoate, or crystallization in the presence of both substrates, resulted in a reaction intermediate, pantoyl adenylate, in the active site [10]. Cocrystallization resulted in both active sites in the dimer being fully occupied with the pantoyl adenylate molecule. However, crystals soaked with ATP and pantoate have one active site (subunit B) fully occupied but the other partially occupied, indicated by their difference in B factors. Subunit A of the dimer also has a lower occupancy of AMPCPP when crystals were soaked in solutions containing AMPCPP. This is due to crystal packing, which makes the flexible wall of the active site cavity ($3_{10}$3-loop-α3'-loop region) ordered in subunit A and thus the active site less accessible for bulky molecules ATP and AMPCPP.

Pantoyl adenylate has extensive binding interactions with the active site residues (**Figure 5**). The molecule is almost linear and fits snugly in the bottom of the active site cavity. Both pantoate and adenylate moieties sit in the same positions in the active site as the pantoate molecule of the pantoate complex and the adenosine group of ATP in the ATP complex, and preserve all strong interactions with protein atoms. The tight binding of pantoyl adenylate stabilizes this highly reactive intermediate. One important feature of the pantoyl adenylate complex is that the disordered loop in subunit B becomes ordered, in both cocrystallization and soaking the apo-enzyme crystals with both ATP and pantoate. This suggests that the flexible wall can function as a gate, opening for substrates, especially the bulky ATP molecule, to bind and closing the active site cavity to protect the highly reactive pantoyl adenylate.

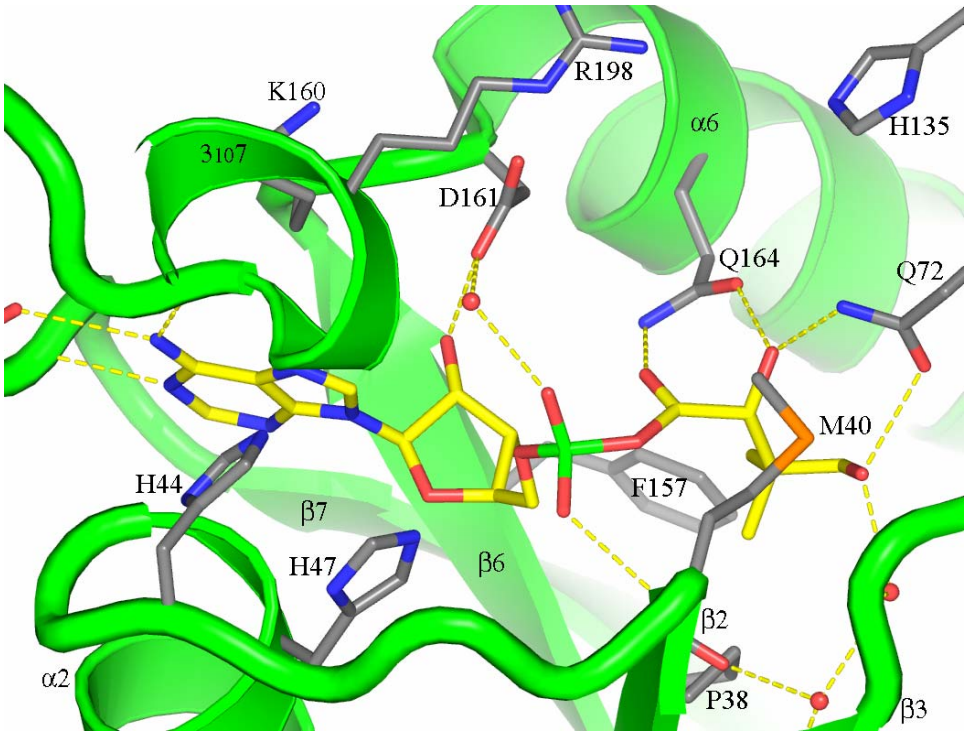



**Figure 5**. Binding interactions of the pantoyl adenylate intermediate. The intermediate binds tightly at the bottom of the active site cavity with many hydrogen bonds and hydrophobic interactions. Hydrogen bonds to the O3* of ribose and some water mediate hydrogen bonds are not shown for clarity.

### 4.7 Binding Interactions of Beta-Alanine

Cocrystallization and soaking crystals with β-alanine [10] failed to yield a complex with this substrate, suggesting that the binding site for β-alanine exists only after pantoyl adenylate is formed in the active site. Based on the structure of the pantoyl adenylate complex, it is likely that the phosphate group of pantoyl adenylate serves as an anchor for the initial binding of β-alanine by offering hydrogen-bonding partners and/or favorable charge-charge interactions. Furthermore, an AMP molecule bound at the active site may also assist the binding of β-alanine from

its phosphate group. Therefore, we soaked crystals of PS in a solution containing both β-alanine and AMP, and we obtained a crystal structure of the PS enzyme in complex with β-alanine and AMP from the $P2_1$ crystals [11].

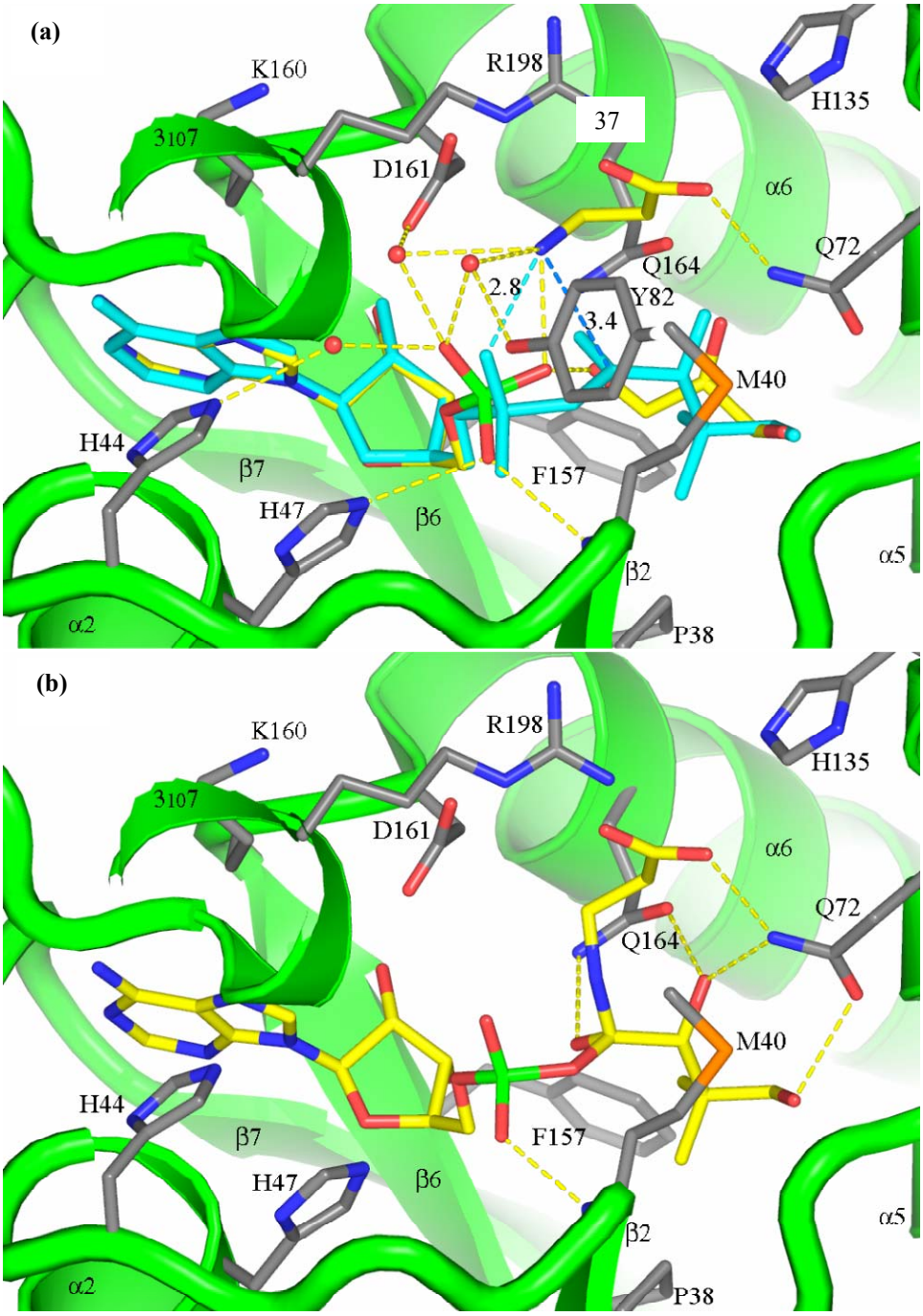



**Figure 6**. **(a)** Structural superposition of the active site of the pantoyl adenylate complex with that of the β-alanine/AMP complex. The active site residues align well with an rmsd of Cα less than 0.15 Å. The pantoyl adenylate molecule is colored in cyan. The protein structure shown in the figure is from the β-alanine/AMP complex. Hydrogen bonds to the phosphate group of AMP and to β-alanine are shown in yellow dashed lines. Some water mediated hydrogen bonds are not shown for clarity. The Tyr82 side chain is from a loop in the front that is clipped away to reveal the bound AMP and β-alanine. Based on the structural alignment, the amino group of β-alanine is ~2.8 Å from the phosphate oxygen of pantoyl adenylate, and ~3.4 Å from the carbonyl carbon, shown in blue dashed lines. **(b)** Model of the tetrahedral intermediate structure based on the structural superposition in (A). The Tyr82 side chain is not shown for clarity. A tetrahedral geometry can be obtained by a torsional rotation of the amino nitrogen of β-alanine to move it to ~1.5 Å above the carbonyl carbon. All favorable interactions with protein atoms are preserved, and thus the intermediate is stabilized by the enzyme.

The β-alanine molecule binds in the active site of subunit A in the upper part of the active site cavity, with its amino group near the phosphate group of AMP (**Figure 6a**). The amino group forms hydrogen bonds to the phosphate group of AMP and to two water molecules. Both water molecules are fixed in position through one hydrogen bond to the phosphate group and a second hydrogen bond to the side chain of Asp161 or Tyr82. The carboxyl group of β-alanine has one hydrogen bond to the Nε2 atom of th[illegible]72 side chain. This carboxyl is also close to the side chains of Arg198 and His135, with distances of ~4.0 [illegible] Å, respectively, indicating favorable charge-charge interactions and π-electron interactions. One side of β-alanine faces side chains of Met40 and Tyr82; the other side faces the large cavity of the active site, which is filled with a few ordered water molecules in the crystal structure. Binding of β-alanine is not as tight as that of other substrates and its binding site in the crystal is not fully occupied. However, the binding interactions limit molecules that can bind well and have a nucleophilic attack on the reaction intermediate. Therefore, structural analogs of β-alanine are all poor substrates [8].

In subunit B of the dimer, where residues 74 to 83 (including Tyr82) are disordered, only AMP but no β-alanine was found in the active site [11]. Also in this case, disorder of this active site cavity wall weakens binding interactions for β-alanine because its binding site is exposed to bulk solvent and thus the charge-charge interactions are dampened. Similarly, in the C2 crystal, which has one molecule per asymmetry unit and the flexible loop disordered, only AMP was found in the active site when crystals were soaked in a solution containing both AMP and β-alanine.

Superposition of the structure of the β-alanine/AMP complex with that of the pantoyl adenylate complex (**Figure 6a**) shows that the amino group of β-alanine was positioned above the carbonyl group of pantoyl adenylate such that the nitrogen atom was ~3.4 Å from the carbonyl carbon. The O1α atom of pantoyl adenylate can form a hydrogen bond with the amino group of β-alanine, serving as an anchor for the initial binding of β-alanine. With changes in only one torsion angle, the amino group of β-alanine can move towards the carbonyl carbon for a nucleophilic attack and attain a nearly perfect geometry for a tetrahedral intermediate (**Figure 6b**).

### 4.8 AMP Binding Interactions

As described above, we obtained AMP complexes from both $P2_1$ and C2 crystals by soaking crystals in a solution containing both AMP and β-alanine. Binding interactions of AMP with the active site residues in these complexes are essentially identical. The adenosine group of AMP binds in the same position as that of the pantoyl adenylate, AMPCPP, and ATP: it fits snugly in its binding pocket at the bottom of the active site cavity through hydrophobic and hydrogen bonding interactions. The adenine group is flanked by Gly46 on helix α2 and Lys160 on the loop after β6. Its N1 and N6 atoms have hydrogen bonds to main chain atoms. The hydroxyl groups of ribose form hydrogen bonds with the Asp161 side chain and a few main chain atoms at the bottom of the active site cavity. The phosphate group, however, has torsional flexibility, and it rotates slightly relative to the α-phosphate of ATP, allowing one of its oxygen atoms to form hydrogen bonds to the side chain of His47 and to the amide nitrogen of Met40 simultaneously (**Figure 6a**).

## 5. OVERALL ENZYME-CATALYZED REACTION MECHANISM

Putting all structural data together allows us to have a detailed view of the overall enzyme-catalyzed reaction mechanism. When the active site is empty, the disordered loop opens the active site and thus allows the ATP molecule to come in and bind. Because ATP is relatively bulky and is held rigidly in the active site by many binding interactions, it is likely that ATP binds first. The smaller sized pantoate then comes in, and initiates a nucleophilic attack on the α-phosphate while it binds in its binding pocket [8, 24]. When both ATP and pantoate are present in the same active site, the nucleophilic reaction must occur (**Figure 4a**). Formation of the transient trigonal bypyramidal intermediate involves only the movement of the phosphate atom towards the carboxyl oxygen of pantoate to be coplanar with the three equatorial oxygen atoms (**Figure 4b**). Although transient in nature, this intermediate is stabilized by the enzyme because all the favorable interactions are preserved, and the unfavorable steric hindrance and charge repulsions between ATP and pantoate are eliminated. The trigonal bipyramidal intermediate then dissociates, and simultaneously the pyrophosphate group leaves the active site. The magnesium ion and the positively charged side chains around the β- and γ-phosphate groups draw the negative charges towards the leaving pyrophosphate. His47 is also likely to facilitate the reaction by donating a proton to the leaving pyrophosphate. This leaves a reaction intermediate, pantoyl adenylate in the active site and completes the first half of the enzyme catalyzed reaction. Pantoyl adenylate is stabilized by many strong binding interactions, as we have seen earlier. It is also protected by the closing of the active site when the flexible loop becomes ordered. Pantoyl adenylate is stable only when tightly bound in the enzyme active site. In solution, it decomposes rapidly to give

pantoyl lactone and AMP [9]. The dissociation of the trigonal bipyramidal intermediate necessitates the simultaneous leaving of the pyrophosphate, because there would be unfavorable charge and steric repulsions once pyrophosphate is formed. This reaction, however, is reversible, as was demonstrated by positional isotope exchange experiments that $^{18}O$ labels scrambled when incubating [$\beta\gamma$-$^{18}O_6$]-ATP and pantoate with the MTB PS enzyme [24].

The presence of pantoyl adenylate intermediate and the ordering of the flexible wall create the binding site for β-alanine, the last substrate. The β-alanine molecule is relatively small, capable of passing through the small opening at the top of the active site cavity. The phosphate group of pantoyl adenylate can have a hydrogen bond to the amino group of β-alanine, thus serving as an initial anchor for binding β-alanine. The carboxylate group of β-alanine has favorable interactions with the side chains of Arg198 and His135 and a hydrogen bond to Gln72. There is a shift of the Arg198 side chain of ~1.8 Å towards the carboxylate group of β-alanine relative to that in the pantoyl adenylate complex. This binding position for β-alanine puts its amino group at a good position for a nucleophilic attack on the carbonyl carbon of the pantoyl group. The distance of the amino nitrogen is only ~3.4 Å from the carboxyl carbon of pantoyl adenylate. A model of the tetrahedral intermediate can be readily prepared from the crystal structures of the pantoyl adenylate complex and the β-alanine/AMP complex (**Figure 6b**). Formation of the tetrahedral intermediate requires only a torsional rotation of the amino nitrogen atom of β-alanine toward the carbonyl carbon of pantoyl adenylate. Similar to that of the trigonal bypyramidal intermediate, formation of the tetrahedral intermediate preserves all favorable binding interactions of active site residues with pantoyl adenylate and β-alanine. Therefore, the intermediate is stabilized by the enzyme. Dissociation of the tetrahedral intermediate then forms pantothenate and AMP. The planarity of the peptide bond of pantothenate causes significant rearrangement of the molecule, and several hydrogen bonds to active site residues are broken. In addition, the newly formed pantothenate has steric clashes and charge repulsions with AMP and some active site residues. Therefore, pantothenate must leave the active site once it is formed. The AMP molecule then diffuses away to regenerate the enzyme active site for next cycle of catalysis. Pantothenate has very low binding affinity to the enzyme. We have tried soaking and growing crystals in the presence of pantothenate, but we failed to obtain a complex with this compound. This is consistent with the finding that pantothenate is a poor inhibitor of the enzyme [8]. The AMP molecule, on the other hand, has good binding interactions in the active site. However, it typically has a low concentration inside cells, and thus can easily diffuse out of the active site. Moreover, ATP has many favorable interactions in the active site of the enzyme, and it can readily displace AMP.

## 6. REFERENCES


[1] S. Jackowski. Escherichia coli and Salmonella typhimurium: Cellular and molecular biology, American Society for Microbiology, Washington, DC 1996, pp. 687-694.

[2] W.K. Maas, Proc Int Congr Biochem 11 (1960) 161-168.

[3] V.K. Sambandamurthy, X. Wang, B. Chen, R.G. Russell, S. Derrick, F.M. Collins, S.L. Morris and W.R. Jacobs, Jr., Nat Med 8 (2002) 1171-4.

[4] W.K. Merkel and B.P. Nichols, FEMS Microbiol Lett 143 (1996) 247-52.

[5] F. von Delft, A. Lewendon, V. Dhanaraj, T.L. Blundell, C. Abell and A.G. Smith, Structure (Camb) 9 (2001) 439-50.

[6] U. Genschel, C.A. Powell, C. Abell and A.G. Smith, Biochem J 341 ( Pt 3) (1999) 669-78.

[7] A. Perez-Espinosa, T. Roldan-Arjona and M. Ruiz-Rubio, Mol Genet Genomics 265 (2001) 922-9.

[8] R. Zheng and J.S. Blanchard, Biochemistry 40 (2001) 12904-12.

[9] T. Wieland, W. Loewe, A. Kreiling and G. Pfleiderer, Biochem Z 339 (1963) 1-7.

[10] S. Wang and D. Eisenberg, Protein Sci 12 (2003) 1097-108.

[11] S. Wang and D. Eisenberg, Biochemistry (2005) In press.

[12] Z. Otwinowski and W. Minor, Methods Enzymol 276 (1996) 307-326.

[13] J. Navaza, Acta Crystallogr A 50 (1994) 157-163.

[14] A.T. Brunger, P.D. Adams, G.M. Clore, W.L. DeLano, P. Gros, R.W. Grosse-Kunstleve, J.S. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson and G.L. Warren, Acta Crystallogr D Biol Crystallogr 54 ( Pt 5) (1998) 905-21.

[15] G.N. Murshudov, A.A. Vagin, A. Lebedev, K.S. Wilson and E.J. Dodson, Acta Crystallogr D Biol Crystallogr 55 ( Pt 1) (1999) 247-55.

[16] T.A. Jones, J.Y. Zou, S.W. Cowan and Kjeldgaard, Acta Crystallogr A 47 ( Pt 2) (1991) 110-9.

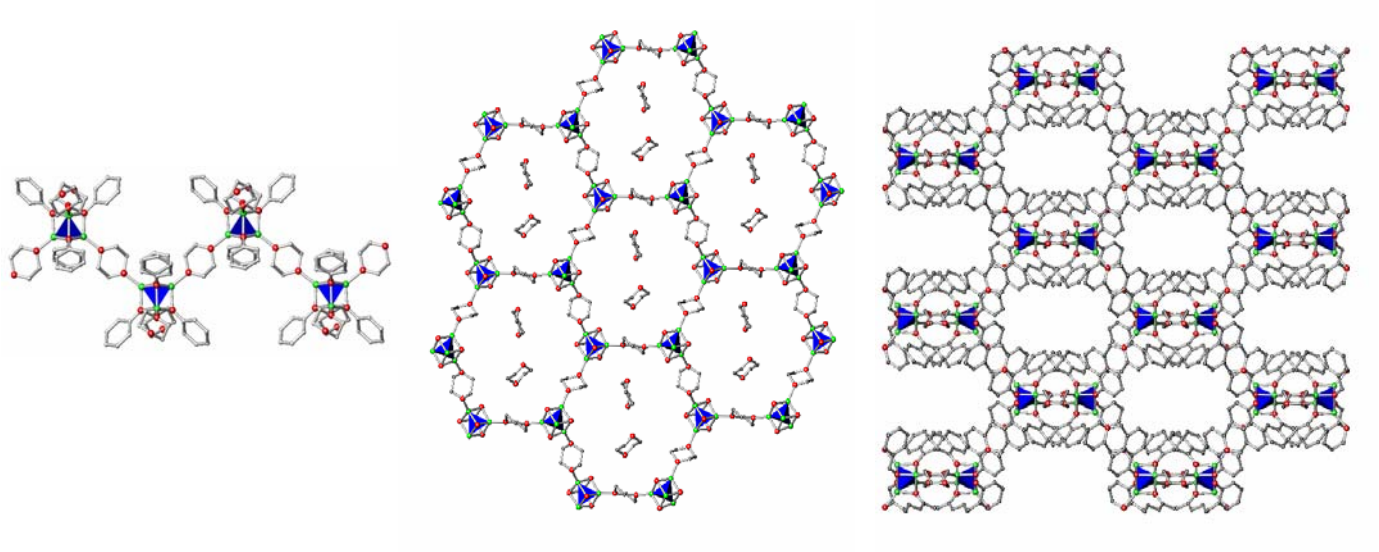
[17] P.J. Kraulis, Journal of Applied Crystallography 24 (1991) 946-950.

[18] E.A. Merritt and M.E. Murphy, Acta Crystallogr D Biol Crystallogr 50 (1994) 869-73.

[19] D.G. Barker and G. Winter, FEBS Lett 145 (1982) 191-3.
[20] W.L. DeLano, The PyMOL molecular graphics system, DeLano Scientific, San Carlos, CA, 2002.
[21] P. Bork, L. Holm, E.V. Koonin and C. Sander, Proteins 22 (1995) 259-66.
[22] R. Zheng, T.K. Dam, C.F. Brewer and J.S. Blanchard, Biochemistry 43 (2004) 7171-8.
[23] C. Hountondji, P. Dessen and S. Blanquet, Bio 40 e 68 (1986) 1071-8.
[24] L. Williams, R. Zheng, J.S. Blanchard and F.M. Raushel, Biochemistry 42 (2003) 5108-13.

*Proceedings from the Symposium*

# Margaret C. Etter Early Career Award



## ETTER TRANSACTIONS OF THE AMERICAN CRYSTALLOGRAPHIC ASSOCIATION

Edited by Frances Jurnak

VOLUME 1, 2005

*Proceedings from the Symposium*

# Margaret C. Etter Early Career Award

ACA ETTER TRANSACTIONS
VOLUME 1

Walt Disney World, Orlando, Florida
May 28 – June 2, 2005

**Editor**

Frances Jurnak
University of California, Irvine
Tel: 949-824-6623
Fax: 949-824-8540
jurnak@uci.edu

# BIS-CHELATION AND ANION EFFECTS INVOLVING A MOLECULE CONSTRUCTED IN THE ORGANIC SOLID STATE USING MOLECULAR TEMPLATES

Tamara D. Hamilton, Giannis S. Papaefstathiou and Leonard R. MacGillivray
*University of Iowa Department of Chemistry, Iowa City, IA 52242*

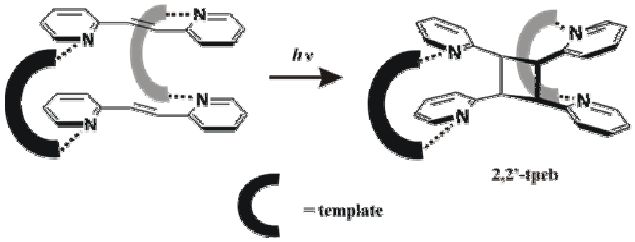
## 1. ABSTRACT

A molecule derived from a template-directed solid-state synthesis, *rctt*-tetrakis(2-pyridyl)cyclobutane (2,2'-tpcb), serves as a bis-chelating ligand upon reaction with $Cu(BF_4)_2{\cdot}H_2O$ and $NEt_4PF_6$ to produce the dinuclear complex $[Cu_2(OH)_2(H_2O)_4(\mu\text{-}2,2'\text{-tpcb})][PF_6]_2$ (**1**).

## 2. INTRODUCTION

Transition-metal-ion complexes involving seven-membered chelation rings have emerged as important for applications in catalysis and biology. Specifically, N-$C_4$-N bridged metallocycles [1] have found applications as catalysts [2] and models of biological compounds [3], as well as antibacterial [4] and antitumor [5] agents. In terms of catalysis, the application of chelating N-donor ligands - as compared to analogous polyphosphines - has begun to attract attention, especially for the activation of C-H bonds [6]. Thus, ligands that provide more than one pyridyl group in close proximity are being studied as preorganized metal chelation units [6b]. In addition, there is increasing interest in the development of novel non-metallocene catalysts for olefin polymerization for which complexes involving N-$C_4$-N bridged chelate rings have been effective [2c]. Therefore, studies that involve N-$C_4$-N ligands with a potential to offer seven-membered chelation are of much current interest.

The ability to position molecules in an orientation appropriate for reaction in the organic solid state using molecular templates that operate *via* hydrogen bonds has led to the quantitative and gram-scale construction of molecules difficult to obtain from solution [7,8]. In particular, we have shown that templates based on resorcinol can orient olefins [*e.g. trans*-1,2-bis(2-pyridyl)ethylene) (2,2'-bpe)] in the solid state in positions suitable for intermolecular [2+2] photodimerizations. The suitability of the products as ligands for coordination chemistry is evidenced by the presence of pyridine rings that emanate from central cyclobutane rings.




**Scheme 1.** Template-directed solid-state synthesis of 2,2'-tpcb from 2,2'-bpe.

As part of an ongoing study to elucidate anion effects involving transition-metal-ion complexes of *rctt*-tetrakis(2-pyridyl)cyclobutane (2,2'-tpcb), we wish to report here the synthesis and structure determination of the dinuclear Cu(II) complex $[Cu_2(OH)_2(H_2O)_4(\mu\text{-}2,2'\text{-tpcb})][PF_6]_2$ (**1**). In line with our previous study [7], the tetrapyridine acts as a bis-chelating bridge with two Cu(II) centers, giving rise to two seven-membered metallocycles. In contrast to our previous work, the $PF_6^-$ counter ions are non-coordinating and, thus, provide an opportunity for each Cu(II) ion to interact with additional donor units. In the case of **1**, each Cu(II) center is coordinated by two water molecules and a single hydroxide anion, each of which participates in hydrogen bonds with the $PF_6^-$ anion.

## 3. METHODS

Complex **1** was prepared by combining methanolic solutions of 2,2'-tpcb (0.025 g, 0.068 mmol) [8a] and $Cu(BF_4)_2{\cdot}H_2O$ (0.032 g, 0.137 mmol) (1:2 ratio). To the resulting clear blue solution was added a methanolic solution of $NEt_4PF_6$ (0.037 g, 0.137 mmol) (25 mL methanol total). A 5 mL aliquot of the latter solution was then

layered with 10 mL benzene in a test tube and allowed to sit for a period of approximately one week. The solution produced blue single crystals (yield: 0.015 g, 29 %) suitable for X-ray analysis [9].

## 4. RESULTS AND DISCUSSION

Compound **1** crystallizes in the space group $P2_1/n$. An ORTEP perspective of the asymmetric unit of **1** is shown in **Figure 1a**. The asymmetric unit consists of one Cu(II) ion, one-half molecule of 2,2'-tpcb, two water molecules, one hydroxide ion, and one $PF_6^-$. As shown in **Figure 1b**, the dinuclear complex, which sits around a crystallographic center of inversion, contains one molecule of 2,2'-tpcb that bridges two Cu(II) centers [d(Cu1···Cu1a) 7.47 Å]. The Cu(II) ions of the complex adopt an *anti* conformation and exhibit a bite angle of 88.8° [10]. Similar to $[Cu_2(NO_3)_4(\mu\text{-}2,2'\text{-tpcb})]$ and $[Cu_2(\mu_2\text{-}SO_4)_2(\mu\text{ -}2,2'\text{-tpcb})(H_2O)_2]_\infty$, each seven-membered chelation ring adopts a boat conformation. The boat conformation also compares favorably to the chelation complexes of *bis*(2-pyridyl)ethane with Pt(II) [11,12] and 1,2-bis(6-methylpyridin-2-yl)ethane with Pd(II) [13]. The coordination geometry around each Cu(II) center of **1** is square pyramidal. Two pyridyl units of 2,2'-tpcb and two water molecules (O1 and O2) form the base of the pyramid while a hydroxide ion (O3) occupies the apical position. The metal-ligand distances around the base of the pyramid range from 1.96-2.02 Å, while the Jahn-Teller distorted apical distance is 2.21 Å (Table 1).

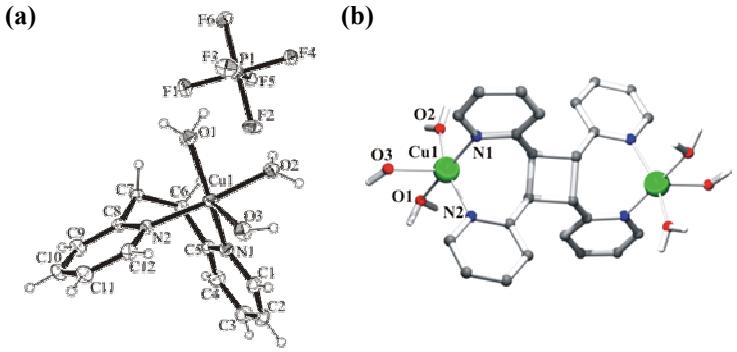



**Figure 1**. **(a)** ORTEP representation of the asymmetric unit of **1**. Displacement ellipsoids are drawn at the 30% probability level, with hydrogen atoms being assigned an arbitrary radius; **(b)** The dinuclear complex **1** (color scheme: green = copper, red = oxygen, blue = nitrogen, grey = carbon). Hydrogen atoms of 2,2'-tpcb have been omitted.

Views of the extended structure of **1** are shown in **Figures 2** and **3**. The complexes have assembled to form layers, parallel to the *ac*-plane, which exhibit an *abab* stacking pattern. The layers are held together by face-to-face π-π forces involving the ligated pyridyl groups (**Figure 2**). The counter $PF_6^-$ ions form layers sandwiched between the layers of Cu(II) complexes. The anions participate in extensive O-H···F hydrogen bonds that involve the two water molecules (O1, O2) and hydroxide ion (O3). Specifically, each $PF_6^-$ anion participates in a total of six O-H···F hydrogen bonds (**Figure 3a**) [d(O1···F3) 2.750(5) Å; d(O1···F4) 2.626(5) Å; d(O2···F2) 2.587(5) Å; d(O2···F6) 2.600(5) Å; d(O3···F1) 2.869(5) Å; d(O3···F5) 2.665(5) Å] (**Figure 3b**) with the ligands.

**TABLE 1**

Selected Bond Lengths (Å) and Angles (°) for **1**

| | | | |
|---|---|---|---|
| Cu1-O1 | 1.987(4) | O1-Cu1-O2 | 84.7(2) |
| Cu1-O2 | 1.963(4) | O1-Cu1-O3 | 99.7(2) |
| Cu1-O3 | 2.214(4) | O1-Cu1-N2 | 97.5(1) |
| Cu1-N1 | 2.020(4) | O2-Cu1-O3 | 83.7(2) |
| Cu1-N2 | 1.986(4) | O2-Cu1-N1 | 92.8(2) |
| | | O3-Cu1-N1 | 101.7(2) |
| | | O3-Cu1-N2 | 100.5(1) |
| | | N1-Cu1-N2 | 88.8(1) |

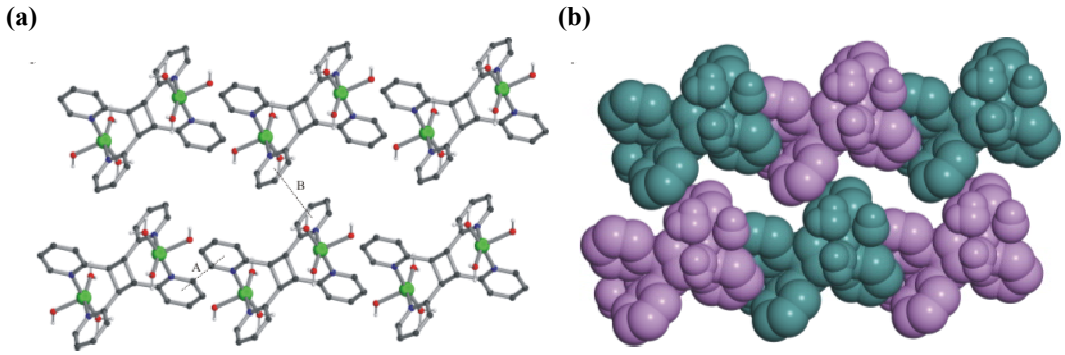



**Figure 2.** Extended structure of **1** showing the π-stacking of pyridine rings viewed in the *ac* plane: **(a)** ball-and-stick representation ($d_A$ = 3.697(5); $d_B$ = 3.874(5) Å) (color scheme: green = copper, red = oxygen, blue = nitrogen, grey = carbon) and **(b)** space-filling representation (hydrogen atoms of 2,2'-tpcb omitted for clarity).

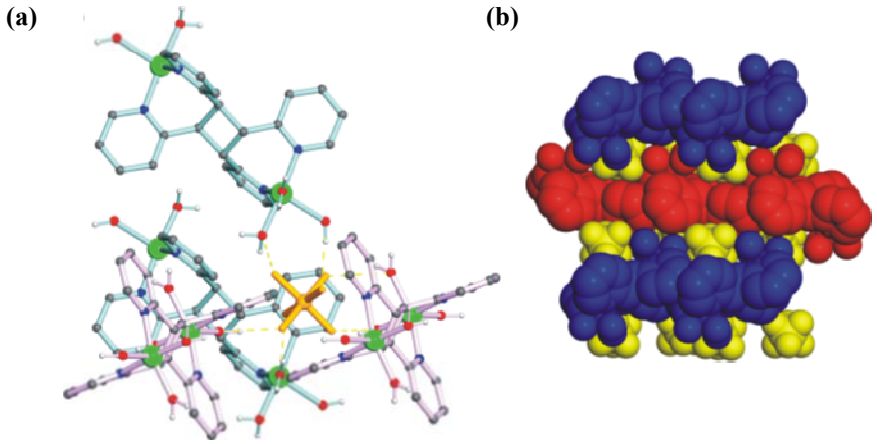



**Figure 3.** **(a)** Hydrogen bonding array involving $PF_6^-$ anion (gold) in **1**. Each F-atom is involved in one hydrogen bond (yellow) to a coordinated water or hydroxide molecule of a neighboring assembly (color scheme: green = copper, red = oxygen, blue = nitrogen, grey = carbon, white = hydrogen) (hydrogen atoms of 2,2'-tpcb omitted for clarity) and **(b)** space-filling representation of **1**, showing *abab* packing of complex layers alternating with $PF_6^-$ anions. Complexes are shown in red or blue, while $PF_6^-$ anions are shown in yellow.

## 5. CONCLUSION

In conclusion, template-directed solid-state organic synthesis has led to the construction of a tetrapyridyl ligand that forms the dinuclear coordination complex **1** with Cu(II) ions. The complex incorporates two seven-membered chelation rings [7]. 2,2'-tpcb is the first ligand to sustain seven-membered chelation in a series of coordination complexes. The non-coordinating behavior of the $PF_6^-$ counter-ion provides open sites on each metal, which are occupied by water molecules and hydroxide ions.

## 6. REFERENCES

[1] a) McCullough, F. Jr.; Bailar, J. C. Jr. *J. Am. Chem. Soc.* **1956**, *78*, 714-716;
b) Habu, T.; Bailar, J. C. Jr. *J. Am. Chem. Soc.* **1966**, *88*, 1128-1130.

[2] a) Gibson, V. C.; Kimberley, B. S.; White, A. J. P.; Williams, D. J.; Howard, P. *Chem. Commun.* **1998**, 313-314;
b) Kretzschmar, E. A.; Kipke, J.; Sundermeyer, J. *Chem. Commun.* **1999**, 2381-2382;
c) Siemeling, U.; Kolling, L.; Stammler, A.; Stammler, H. G.; Kaminski, E.; Fink, G., *Chem. Commun.* **2000**, 1177-1178;
d) Bacchi, A.; Pelizzi, G.; Minardi, G.; Pistuddi, A. M.; Solinas, C.; Chelucci, G. *Trans. Met. Chem.* **2002**, *27*, 274-278;
e) Minardi, G.; Mura, E.; Pistuddi, A. M.; Solinas, C.; Bacchi, A.; Pelizzi, C.; Pelizzi, G.; Chelucci, G *Trans. Met. Chem.* **1999**, *24*, 481-485;

f) Zhu, H. B.; Dai, Z.-Y.; Huang, W.; Cui, K.; Gou, S.-H.; Zhu, C.-J. *Polyhedron* **2004**, *23*, 1131-1137;
g) Bandoli, G.; Gerber, T. I. A.; Jacobs, R.; du Preez, J. G. H. *Inorg. Chem.*, **1994**, *33*, 178-179.

[3] a) Martin, J. W. L.; Timmons, J. H.; Martell, A. E.; Rudolf, P.; Clearfield, A. *Inorg. Chem.* **1981**, *20*, 814-821;
b) Choo P. L.; Mulichak, A. M.; Jones Jr., R. W.; Bacon, J. W.; Pett, V. B.; Zacharias, D. E. *Inorg. Chim. Acta* **1990**, *171*, 183-192;
c) Ball, R. G.; Trotter, J. *Can. J. Chem.* **1979**, *57*, 1368-1371;
d) McFarland, C. A.; Gross, S. A.; Winfield, J. S.; Pett, V. B. *Inorg. Chim. Acta*, **1994**, *221*, 35-41;
e) Balamurugan, R.; Palaniandavar, M.; Gopalan, R. S. *Inorg. Chem.* **2001**, *40*, 2246-2255.

[4] Nandi, A. K.; Chaudhuri, S.; Mazumdar, S. K. Ghosh, S. *Inorg. Chim. Acta* **1984** *92*, 235-240.

[5] a) Khokar, A. R.; Shamsuddin, S.; Xu, Q. *Inorg. Chim. Acta* **1994**, *219*, 193-197;
b) Shamsuddin, S.; van Haal, J. W.; Stark, J. L.; Whitmore, K. H.; Khokhar, A. R. *Inorg. Chem.* **1997**, *36*, 5969-5971;
c) Shamsuddin, S.; Santillan, C. C.; Stark, J. L.; Whitmire, K. H.; Siddik, Z. H.; Khokhar, A. R. *J. Inorg. Biochem.* **1998**, *71*, 29-35;
d) Hoeschele, J. D.; Showalter, H. D. H.; Kraker, A. J.; Elliott, W. L.; Roberts, B. J.; Kampf, J. W. *J. Med. Chem.* **1994**, *37*, 2630-2636.

[6] a) Vedernikov, A. N.; Huffman, J. C.; Caulton, K. G. *Inorg. Chem.* **2002**, *41*, 6244-6248;
b) Vedernikov, A. N.; Huffman, J. C.; Caulton, K. G. *Inorg. Chem.* **2002**, *41*, 6867-6874; Vedernikov, A. N.; Pink, M.; Caulton, K. G. *Inorg. Chem.* **2004**, *43*, 3642-3646.

[7] Hamilton, T. D.; Papaefstathiou, G. S.; MacGillivray, L. R. *Cryst. Eng. Comm.* **2002**, *4*, 223-226.

[8] a) Papaefstathiou, G. S.; Kipp, A. J.; MacGillivray, L. R. *Chem. Commun.* **2001**, 2462-2463;
b) MacGillivray, L. R.; Reid, J. L.; Ripmeester, J. A. *J. Am. Chem. Soc.* **2000**, *122*, 7817-7818;
c) Hamilton, T. D.; Papaefstathiou, G. S.; MacGillivray, L. R. *J. Am. Chem. Soc.* **2002**, *124*, 11606-11607.;
d) Varshney, D. B.; Papaefstathiou, G. S.; MacGillivray, L. R. *Chem. Commun.* **2002**, 1964-1965;
e) Gao, X.; Friščić, T.; MacGillivray, L. R. *Angew. Chem. Int. Ed.* **2004**, *43*, 232-236.

[9] Crystal data for **1**: monoclinic, $P2_1/n$, Z = 4, a = 9.716(5), b = 16.586(5), c = 9.939(5) Å, $\beta$ = 98.270(5) º, V = 1585.0(12) Å$^3$, 2738 unique reflections, $R$ = 0.0492 [for 2122 reflections with $I \geq 2\sigma(I)$], $R$ = 0.0742 and w$R^2$ = 0.1408 (all data). Crystallographic Data has been deposited with the Cambridge Crystallographic Data Centre and has been allocated with deposition number CCDC 291578.

[10] The bite angle is the angle between chelating ligand donor atoms (L-M-L) in a metal complex.

[11] McFarlane, A.; Lusty, J. R.; Fiol, J. J.; Terron, A.; Molins, E.; Miravitlles, C.; Moreno, V. *Z. Naturforsch., Teil B*, **1994**, *49*, 844-848.

[12] a) Marcelis, A. T. M.; Korte, H.-J.; Krebs, B.; Reedijk, J. *Inorg. Chem.*, **1982**, *21*, 4059-4063;
b) Kleibohmer, W.; Krebs, B.; Mercelis, A. T. M.; Reedijk, J.; Van der Veer, J. L. *Inorg. Chim. Acta* **1983**, *75*, 45-50.

[13] Baker, A. T.; Crass, J. K.; Kok, G. B.; Orbell, J. D.; Yuriev, E. *Inorg. Chim. Acta* **1993**, *214*, 169-176.

# USE OF ALKALI METAL AGGREGATES IN CONTROLLING NETWORK ASSEMBLY

Dugald J. MacDougall, J. Jacob Morris, Bruce C. Noll and Kenneth W. Henderson*
*Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, Indiana 46556, USA.*

## 1. ABSTRACT

This paper summarizes the first systematic studies into the use of *s*-block aggregates to control the rational assembly of network polymers. Preliminary work using lithium and sodium aryloxide aggregates has established the efficacy of this methodology, and in particular that the control of network architectures can be achieved on a regular basis. These studies also demonstrate that air-stable, robust, open-framework materials may be prepared through use of these systems. Overall, the paper outlines an area that has considerable potential for the preparation of new materials, that advance knowledge in the understanding of the self-assembling process.

## 2. INTRODUCTION

The synthesis and characterization of periodic network assemblies is an area of increasing international academic and industrial interest [1-4]. The attention paid to this topic can be attributed, in part, to the fundamental scientific challenges present in developing rational routes to structurally well-defined solids and also as a consequence of the potential utility of these materials in applications as diverse as catalysis [5], chemical separation [6], optics [7] and electronics [8]. In this regard, two main strategies have been adopted in the formation of frameworks from molecular precursors [9], firstly through the use of hydrogen-bonded organic solids [10] and secondly using metal-organic coordination polymers containing transition elements and ligand spacer molecules [11]. These strategies have also been combined where metal-containing complexes are incorporated within hydrogen-bonded networks [12]. In addition, a notable area of success in controlling the formation of covalently-linked solids has been the use of metal carboxylate clusters as directional components of metal-organic frameworks (MOFs) [13]. However, a prominent omission from the current list of strategies directed towards rational network synthesis is the use of early main group metals to dictate supramolecular structure. This deficit may be explained by the considerable potential problems associated with these species, such as the presence of complex dynamic solution equilibria between aggregated complexes; the spherical non-polarizable nature of the cations; the limited preference for fixed metal geometries, and the lability of the metal-ligand bonding. All of these issues may lead to poor predictability and reproducibility in network synthesis [14-16]. In this context it is unsurprising that researchers have opted for alternative structural scaffolds, and to a large extent have consciously avoided the use of *s*-block metal cations. In this paper we directly address these established perceptions and contest that the early main group elements can be successfully employed to control network assembly by the informed selection of metal, anionic partner, linker molecule and synthetic conditions. More specifically, we outline a novel approach to rational network synthesis utilizing pre-assembled *s*-block molecular aggregates as secondary building units (SBUs) to control supramolecular structure. Although the use of SBUs is quickly emerging as a successful tactic in rationalizing and controlling the formation of various types of extended framework architectures, the use of *s*-block SBUs has not been investigated [17]. In this paper we detail our preliminary investigations in this area, which firmly establish the efficacy of this methodology. We will outline an unprecedented systematic series of one-, two- and three-dimensional polymers constructed from *s*-block metal aggregates. A subset of these networks displays intriguing host-guest chemistry, including the preparation of air-stable, robust, open-framework materials. We have also developed a model for network prediction, which will be an essential element in our future studies this area.

### 2.1 Initial Observations and Proof-of-Concept Preliminary Studies

During the course of our investigations into the structure and function of $\alpha$, $\alpha$'-stabilized carbanions we discovered that the lithium salts of many of these species form unusual polymeric materials [18]. We structurally characterized lithiated $\alpha$-cyanophosphonates of the type $[(RO)_2P(O)CHCNLi.THF]$ (R = Et or $Pr^i$) and found that they form 2D sheet structures, as displayed in **Figure 1**.

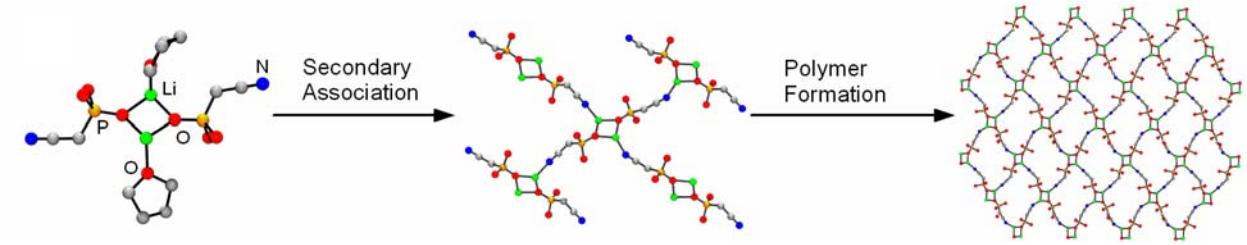



**Figure 1.** Assembly process of lithiated α-cyanophosphonates by linking $Li_2O_2$ dimers.

Intriguingly from the molecular perspective, the sheets can be described as being composed of 'dimeric' $Li_2O_2$ units that are interconnected to four identical 'dimers' via ligation of the nitrile groups. The $Li_2O_2$ motif, where the metals bridge between a pair of phosphoryl units, is entirely consistent with established aggregation patterns for simple lithiated phosphonates [19]. Then, the pendant nitriles act as linear-linkers and bond in an 'interdimer' fashion to another $Li_2O_2$ unit, with the adjacent rings rotated by approximately 90° with respect to one another. Overall, four dimers interconnect through the nitrile bridges to give large $(LiOPCCN)_4$ twenty-four membered rings with small four membered $Li_2O_2$ rings at the corners. This pattern is topologically equivalent to a (4,4) net and can be described as a decorated network where one square-planar vertex has been replaced by a molecular aggregate [20, 21]. The networks obtained are reminiscent of classical structures utilizing square-planar transition metal nodes, such as $Ni^{2+}$, $Zn^{2+}$, $Cu^{2+}$, $Cd^{2+}$, *etc.*, with neutral linear-linker ligands such as 4,4'-bipyridine. In our case the $Li_2O_2$ dimeric aggregates act as square-planar nodes, where the nodal point is located at the center of the ring rather than being an individual atom.

We theorized that it should be possible to form related architectures in a rational and controlled manner using other types of lithiated aggregates. We chose to examine (organo)sulfonylacetonitriles, $RSO_2CH_2CN$, as carbanion sources since simple lithiated sulfonyls are known to form $(SO_2Li)_2$ eight membered ring dimers [22]. Therefore, these complexes should assemble in a similar manner to the cyanophosphonates described above but with larger eight membered rings interconnecting the macromolecular twenty-four membered rings. In this vein we targeted and subsequently successfully prepared and structurally characterized the complex $[MeSO_2CHCNLi.THF]$, [23] and we were pleased to confirm that this complex forms a network with exactly the connectivity that was predicted in advance [24, 25]. **Figure 2** illustrates the formation of this two dimensional sheet that may be described as a basket weave network. Overall, the $(SO_2Li)_2$ eight-membered ring acts as an offset square planar node to give the desired topology.

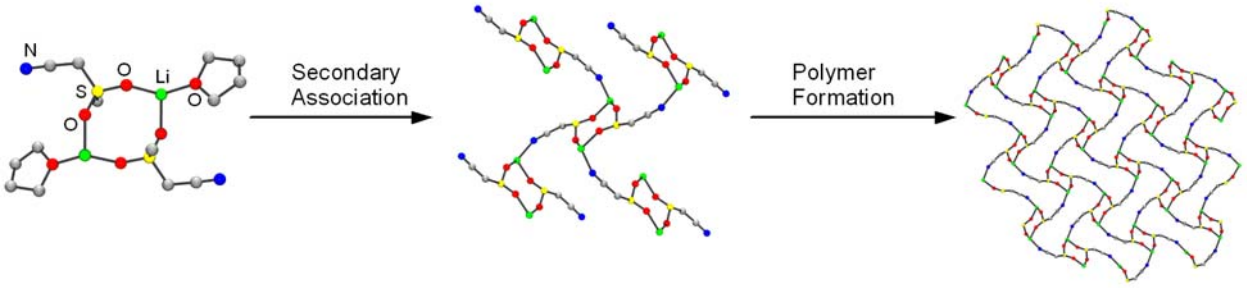



**Figure 2.** Assembly process of lithiated (organo)sulfonylacetonitriles through association of $(SO_2Li)_2$ dimers.

## 3. METHODS

The coordination chemistry involving the early main group elements is an active area of academic study, with the molecular aggregation behavior of numerous classes of metal-ligand combinations now well established [14-16]. With this understanding of molecular aggregation in hand, we aimed to select a series of readily prepared and well-behaved main group aggregates that could then be utilized as SBUs. A highly attractive feature of *s*-block SBUs is the diversity of metal aggregate geometries and compositions available as synthons. Our goal was to utilize the

geometric and chemical information stored in a variety of metal-containing ring and cage compounds to rationally design, and subsequently build, novel two- and three-dimensional network architectures. Specifically, since the location of the metal centers and the anions within these aggregates are known in advance, we could use this information to predict the directions for polymer extension. Suitable aggregates were selected that contain metals with one or more free coordination sites or labile ligands for Lewis base ligation and subsequent polymer growth. In essence the entire aggregate is used as a fixed nodal point (as opposed to the nodes being single spherical metal cations) with their geometries determined by the nature of the rings or cages. This strategy includes utilizing strong 'primary' bonding interactions to assemble molecular aggregates of predictable composition, shape and size, which are then linked via weaker 'secondary' Lewis base-type bonding. The interaggregate association is achieved by the use of neutral polydentate Lewis bases containing at least two donor sites, as opposed to the use of internal donor groups attached to the anionic ligand backbones as described in **Figures 1** and **2**. The use of SBUs in network synthesis is also an appealing approach since they dramatically reduce the number of possible network topologies arising for a given node/linker combination and they may also aid in the preparation of stable porous materials [26].

It is pertinent to recognize that polymer formation is in fact very common for early main group compounds, in particular for *s*-block complexes [27, 28]. Indeed, most preparative chemists have from time to time encountered an insoluble lithiated intermediate during a synthesis. In many instances such solubility problems can be overcome by the use of polar solvents such as THF, DME and TMEDA, which cleave metal-ligand interactions and generally reduce the aggregation state of the metallated species. In turn, the vast majority of work carried out on the structure and bonding of early main group metal complexes has been performed at the molecular level, and the structural elucidation of polymeric materials has mainly been by inference rather than by direct analysis [29]. Although, various types of coordination polymers containing *s*-block metals have previously been characterized, they have usually been prepared either inadvertently or in order to study localized metrical information, rather than deliberately to build specific network architectures [30]. Indeed, while elegant work has been carried out elucidating the subtle factors involved in determining the molecular structures adopted by *s*-block metal complexes, no sustained attempt has been made to rationalize polymer formation in this area.

### 3.1 Common Geometries Available for Alkali Metal Containing SBUs

The use of lithiated SBUs was a reasonable starting point for the study since the understanding of the aggregation behavior for these complexes is by far the most developed within the *s*-block series [14-16]. Our initial phase of the project utilized the abundance of information that is available for lithiated species to choose suitably robust molecular aggregates as building blocks. **Figure 3** displays the core structures of a selection of commonly found aggregation states for *s*-block complexes: ring dimers, ring trimers, tetrameric cubanes, and prismatic hexamers. **Figure 3** also illustrates the relationship between the metal arrangements within the aggregates and their corresponding nodal geometry. The exact arrangement of the metals within each type of aggregate will vary with the identity of its anionic partner. This will lead to a variety of geometric distortions from the 'ideal' nodal configuration. Nevertheless, the basic ring or cage arrangements should remain intact and will therefore control the direction of polymer assembly.

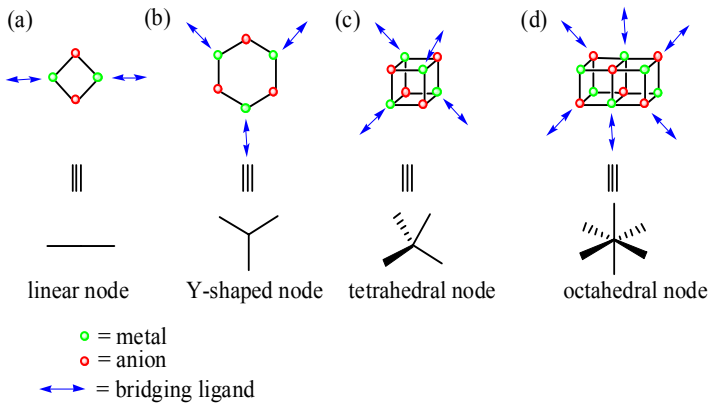



**Figure 3.** Relationship between selected metal aggregates and their related nodal geometries: (a) disolvated ring dimer, (b) trisolvated ring trimer, (c) tetrasolvated tetrameric cubane, and (d) hexasolvated prismatic hexamer.

### 3.2 Lithium Aryloxides as SBUs

It was essential to the success of our approach that the aggregation state of the metallated compound was predictable and remains intact both in solution and in the solid-state while in the presence of the Lewis base linker. An excellent starting point was lithium aryloxides. These complexes contain strong Li-O bonding, and numerous aggregate types are available as SBUs. In addition, they are readily prepared from the parent phenol by direct deprotonation using bases such as organolithium or lithium amide reagents. The aggregate size and coordination environment around the metal center can readily be controlled by the choice of phenol and donor solvent. For example, increasing the steric bulk at the 2,6-positions of an aryloxy ring results in decreasing the aggregation state of the lithiated complex. This is conveniently demonstrated for a series of THF-solvated lithiated aryloxides (ArOLi), shown in **Figure 4**, which form stable ring dimers (Ar = 2,6-$^{t}Bu_2C_6H_3$), ring trimers (Ar = 2,6-$^{i}Pr_2C_6H_3$), tetrameric cubanes (Ar = 2,4,6-$Me_3C_6H_2$), and prismatic hexamers (Ar = $C_6H_5$) [31-33].

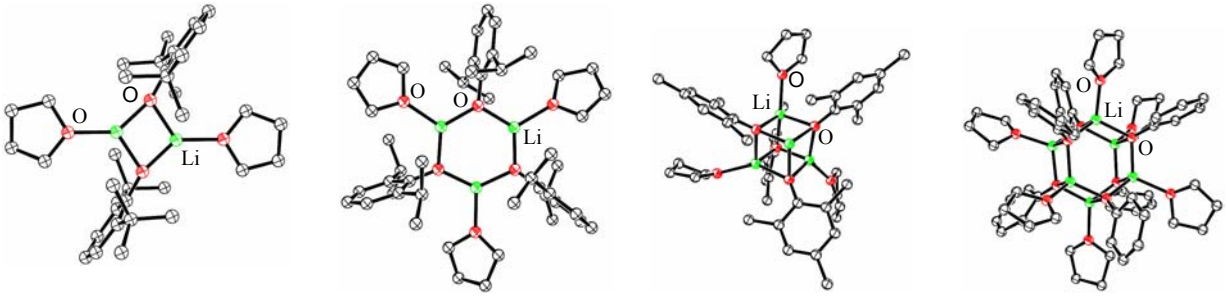



**Figure 4.** Molecular structures of a series of THF-solvated lithium aryloxides, $[ArOLi.THF]_n$.

Each of the metal atoms within these aggregates has one coordination site occupied by a monodentate THF donor molecule. This makes these anionic aryloxide ligands excellent targets as SBUs since replacement of the terminal Lewis bases by didentate linear-linker molecules will result in interconnecting aggregates. There are of course numerous possibilities for both the connectivity and topology of the supramolecular assemblies that may be formed via the interaction of the four aggregate types shown above with a linear-linker such as dioxane. However, it is becoming clear that a relatively few, high-symmetry structures dominate network topologies [26]. The most likely assemblies to be produced in each case are one-dimensional chains from disolvated dimers; hexagonal (6,3) sheets from ring trimers; three-dimensional diamondoid networks using tetrameric cubanes; and cubic lattices using prismatic hexamers.

### 3.3 Choice of Neutral Linker Ligand

Polydentate Lewis bases are required in order to connect neighboring aggregates. The strategy used to achieve this goal was to employ neutral linear-linker ligands containing relatively hard Lewis base donors [34]. A selection of potentially useful linkers for the *s*-block systems are shown in **Figure 5**.

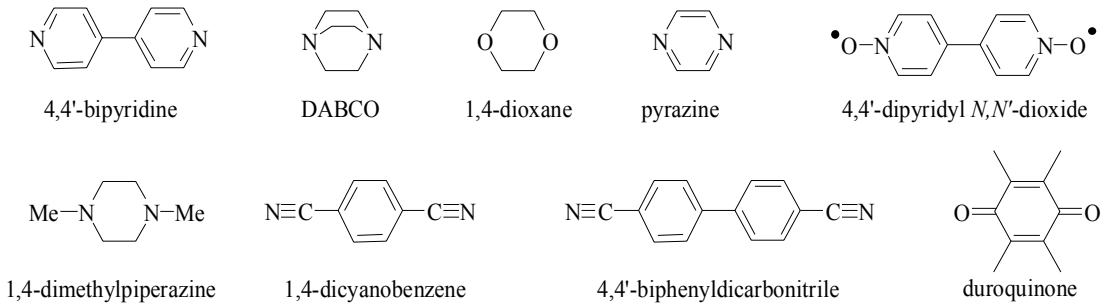



**Figure 5.** Neutral linear-linker molecules suitable for coordination to hard *s*-block metals.

Generally, divergent ligands have been used in lithiated systems simply to enhance the crystallization properties of a metallated molecular substrate in order to study localized bonding or metrical information, rather than to

deliberately build specific network structures [30, 35, 36]. However, in most instances no recognizable lithiated aggregate is present, rather isolated ions with no inherent directional control are linked together. Furthermore, reports on the use of other linear-linking didentate Lewis bases for the *s*-block metals are surprisingly scarce. This deficiency is remarkable considering the enormous importance of such ligands in the assembly of transition metal-containing networks [37]. Nevertheless, ligands of this type should be ideal candidates for coordination to *s*-block metals, in particular lithium, due to their excellent properties as hard Lewis bases. The monodentate base pyridine is a commonly used donor solvent for lithiated complexes with dozens of structurally characterized examples in the Cambridge Structural Database [38], and furthermore several crystal structures of lithium species have been characterized containing the chelating ligand 2,2'-bipyridine [39]. In the present study the linker of choice is 1,4-dioxane as this ligand proved to be convenient as it not only acts as a divergent linker but also as solvent media.

## 4. RESULTS AND DISCUSSION

### 4.1 Use of Tetrameric $Li_4O_4$ Cubanes as SBUs Linked by Dioxane

This early success in rationally designing complex network architectures from very simple constituents spurred us to expand the scope of our studies to include the possibility of linking together neutral, pre-assembled lithiated SBUs using *external* neutral divergent Lewis bases. This is appealing since it immediately opens up a wide variety of well-studied solvated aggregate types for investigation. We targeted the commonly encountered $Li_4O_4$ tetrameric cubanes formed by lithium aryloxides (ArOLi) for use as SBUs [31]. These complexes are excellent SBU candidates since they contain strong Li-O bonding, the metals are held in an approximately tetrahedral arrangement with one coordination site available for ligation, and there is a wide range of substituted phenols available for systematic studies. The didentate donor dioxane was selected as the divergent external linking Lewis base due to its capacity to act as solvent media, its rigidity and its inability to chelate a lithium center.

At the outset, an exploratory computational study was conducted to determine the feasibility of dioxane acting as a bridging ligand between a pair of $Li_4O_4$ cubanes, since this structural pattern had not previously been characterized (**Figure 6**). Semi-empirical PM3 calculations were the first line of analysis due to their computational efficiency [40, 41]. Geometry optimization calculations were performed on the full molecules [(PhOLi.dioxane)$_4$] and [{(PhOLi)$_4$}$_2$.(dioxane)$_7$] rather than using simplified model complexes to ensure an accurate portrayal of subtle steric and electronic effects. These calculations indicated that no notable energy penalty (<0.2 kcal/mol) is incurred on bridging a pair of tetrasolvated cubanes by the donor. This key result was confirmed by further geometry optimization at the *ab initio* HF/6-31G* level of theory (<0.01 kcal/mol) and also by high level single-point density functional theory calculations (B3LYP/6-311G**, <0.04 kcal/mol) [41-44]. These calculations firmly established the viability of polymer formation. In addition, this study demonstrates the ability of lower level calculations to be used as an efficient method for screening suitable aggregate/linker combinations.

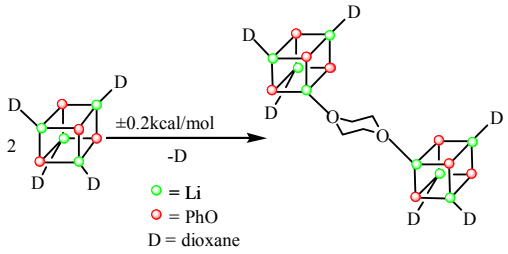



**Figure 6.** Calculated energetics of dioxane bridging a pair of $Li_4O_4$ units.

Appropriate lithium aryloxides were then chosen based on their predilection to form tetrasolvated $Li_4O_4$ cubanes in the presence of monodentate Lewis bases [23]. The required tetrameric aggregation was confirmed by preparing and structurally characterizing (by X-ray diffraction, XRD) the molecular analogues using monodentate donors (THF or pyridine), and also by solution NMR studies of the complexes in dioxane solvent media using the Jackman method of correlating the $^{13}C$ NMR shift positions for the *para*-carbons with aggregation state [25, 45]. Specific structural variants of these ligands were then selected as our understanding of these systems evolved. **Figure 7** highlights the structural characterization, by single-crystal XRD, of the three representative complexes [{(ROLi)$_4$.(dioxane)$_x$}$_\infty$], (a) R=Ph, x=3; (b) R=4-Et-$C_6H_4$, x=2.5; and (c) R=1-naphth, x=2, which typify our findings [46].

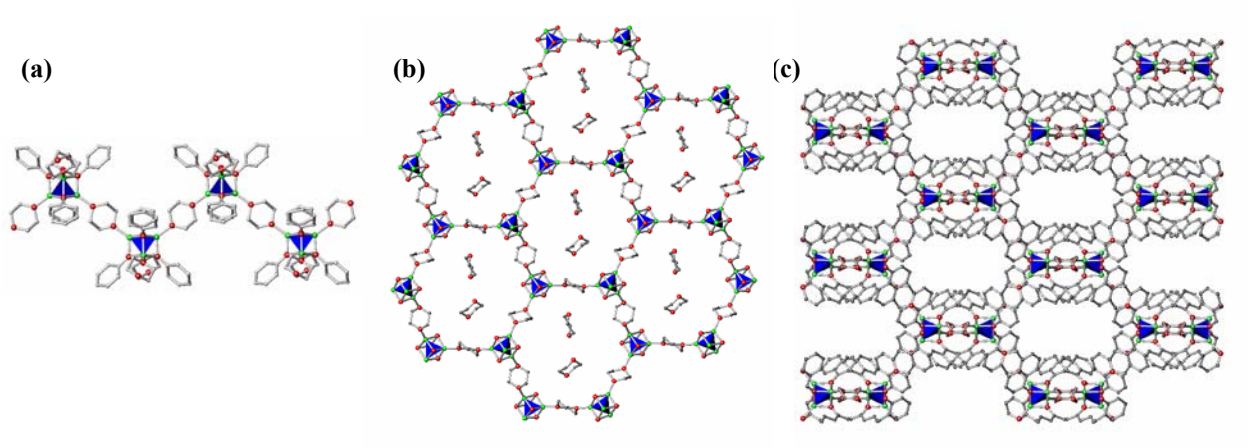



**Figure 7.** Sections of the polymeric structures of the: **(a)** 1D zig-zag chain, **(b)** framework atoms of the 2D hexagonal network with two guest solvent molecules per macrocycle, and **(c)** 3D diamondoid network. Vertices of the blue tetrahedra represent the four lithium centers within each cubane.

The first key point is that each structure is composed of linked $Li_4O_4$ tetrameric cubane units, proving that the desired molecular SBU remains intact on polymerization. Next, the three polymeric types found, 1D zig-zag chains, 2D hexagonal nets, and 3D diamondoid networks are very closely related, simply differing by the number of points of extension from the tetrahedral SBU (two, three and four leading to V-shaped, trigonal and tetrahedral nodes respectively). Moreover we have now fully characterized an extended series of $Li_4O_4$ substituted aryloxide complexes and found that they each form one of these three structural types. Pleasingly, this demonstrates that the SBU approach is indeed successful in limiting the number of network topologies for a given node/linker combination.

The complexes in **Figure 7** illustrate the dramatic effect of relatively small changes in sterics on the macromolecular structure adopted. We deduced that the type of polymer formed can be rationalized in terms of balancing entropic factors with the efficient filling of space. Although our calculations indicate that bridging by dioxane is thermoneutral, there will be a substantial increase in entropy associated with polymer formation due to the gain in translation freedom of the liberated solvent molecules. Next, there will clearly be an optimum length for the rigid anionic ligands of the SBUs in order to efficiently fill the hexameric macrocycles of the 2D sheets. If the ligands are too short this will result in energetically unfavorable large voids at the center of the rings. Conversely, if the ligands are too long the 2D arrangement will be destabilized by transannular interactions [47]. This analysis is consistent with the 1D chain structure observed for the small $PhO^-$ anions. Extension of the ligand at the *para*-position partially circumvents this problem and the longer 4-Et-$C_6H_4O^-$ anions, in combination with the guest dioxanes, fill the hexameric macrocycles with reasonable efficiency. However, examination of the extended crystal-packing diagram of this complex shows interdigitation between neighboring 2D layers, suggesting that significant space still remains within the sheet. In order for a diamondoid structure to be formed it is necessary to occupy 3D adamantanoid cavities. This proves possible for 1-naphthOLi by lateral extension to the aromatic ring, leading to excellent localized space filling in the 'corners' of the adamantanoid cavities by interlocking of the naphthyl units [47]. Space filling by interpenetration is precluded due to the $(1\text{-naphthOLi})_4$ cubanes being ~14.2 Å in diameter whereas the adamantanoid cavities have dimensions of only 7.3 x 7.6 x 16.6 Å. The open structure found for this complex supports the conjecture that the use of SBUs promotes the formation of open-framework materials. The viability and scope of our rationalization was tested through targeting the structural characterization of specific substituted aryloxides. This proved successful, such that when the *para*-substituent is small (H, Me, F or Cl) 1D chains are produced, minimal extension at this position (Et, OMe, $^i$Pr) leads to the formation of 2D hexagonal sheets, whereas lateral substitution of the aromatic (Ar = 1-naphth or 2,4,6-$Me_3C_6H_2$) gives 3D diamondoid networks. In all cases the homogeneity of the samples was confirmed by powder XRD of the bulk solids.

The potential for porosity is an exciting feature of the diamondoid structures outlined above and our preliminary investigations indicate that the guest molecules within $[\{(1\text{-naphthOLi})_4.(\text{dioxane})_2\}_\infty]$ (**Figure 7c**) can be removed under reduced pressure without disruption to the framework. Solvent loss was monitored by $^1$H NMR spectroscopy through complete dissolution of samples in $d_6$-DMSO, followed by integration of the anion and dioxane signals.

Initially, three enclatherated solvent molecules per $Li_4O_4$ unit are present but after 24 hours of evacuation ($1.5 \times 10^{-3}$ Torr and 30°C) the non-framework dioxane is entirely removed. Calculations indicate that 34.8% of the total volume within this structure is potential solvent space.[77] Comparison of the calculated, as prepared, and evacuated powder XRD patterns show only minor variations, demonstrating that the integrity of the framework is maintained on solvent loss. In addition, the macroscopic morphology of the crystals is retained after evacuation (perfect octahedra) allowing determination of the unit cell parameters by single-crystal XRD, and pleasingly they match those of the original material (a full structure determination has not yet been completed due to the weak diffraction of the crystals). Furthermore, these crystals were found to be completely air-stable over several weeks. Unit cell checks confirmed that the crystallinity is retained and IR spectroscopy showed no detectable decomposition to lithium hydroxide.

### 4.2 Use of Hexameric $Na_6O_6$ Prismatic Cages as SBUs Linked by Dioxane

Many complexes of the heavier Group 1 metals are appealing as SBUs. Of the heavier alkali metals, the alkoxides and aryloxides of both sodium and potassium have been reasonably well studied [48, 49]. A potential drawback of moving from lithium to the heavier Group 1 metals is a significant weakening of the metal-ligand bond strength, which may in turn lead to increased aggregate lability and a decrease in the predictability of the nature of the SBU. To test the feasibility of using these metal aggregates we targeted solvated $Na_6O_6$ hexameric aggregates as potential sodium based SBUs since this structural type has been characterized in the solid state for molecular derivatives, *e.g.* $[(PhONa.THF)_6]$ [50]. We theorized that the triple stack of dimers (face-shared cubane) aggregates should act as octahedral SBUs and lead to the formation of 3D cubic networks. Building on our work on the lithiated systems, we utilized 4-substituted aryloxides as convenient probes to test the effect of sterics on the supramolecular arrangements obtained. The set of *para*-halide substituted derivatives are particularly useful in illustrating the effect of sterics on network structure. The complexes $[(4\text{-R-}C_6H_4ONa)_6.(dioxane)_x]_\infty$, where (a) R = I, x = 5 form 1D linear chains, (b) R = Br, x = 4 from (4,4) 2D square nets, and (c) R = F and x = 3 form 3D cubic lattices, **Figure 8** [51]. This remarkable set of complexes again demonstrates that the desired hexameric $Na_6O_6$ SBU is robust and remains intact within each network. Also, the type of network can again be rationalized in terms of the size of the substituent at the *para* position, with only the fluorine group being small enough to fit into the available volume within the cubic cavities. Moreover, we have again confirmed that network formation using the $Na_6O_6$ SBUs is limited to these three structural types, *i.e.* structural characterization of the dioxane solvated *para*-aryloxides R-$C_6H_4ONa$: R = Cl, I or Ph, give 1D linear chains, R = Br, H, Me, $^i$Pr, $^t$Bu or $^t$Pn give 2D (4,4) nets and R = F or Et give 3D cubic networks. These results show that even the heavier *s*-block elements may be used as robust SBUs.

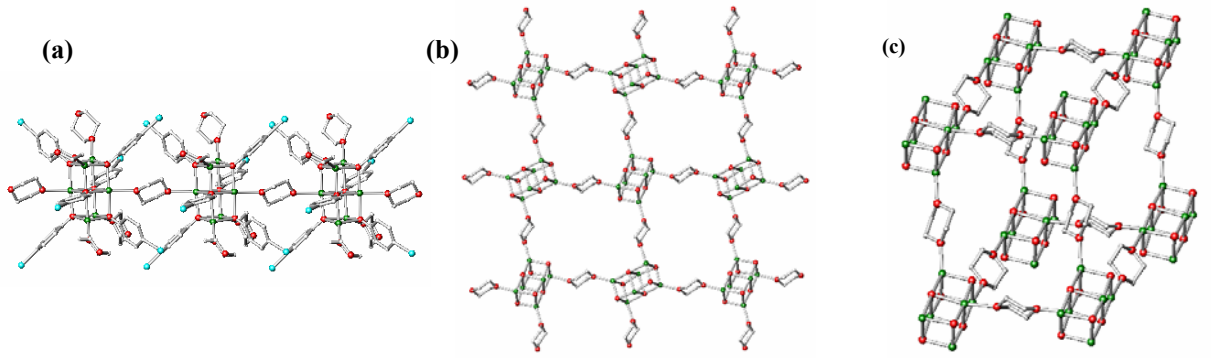



**Figure 8.** Sections of the polymeric structures of the: (a) 1D linear chain, **(b)** 2D square net (framework atoms only), and **(c)** 3D cubic lattice (framework atoms showing a single cubic unit), formed from triple-stack $Na_6O_6$ SBUs bridged by dioxane.

## 5. CONCLUSIONS

To summarize, we have demonstrated that both lithiated and sodiated pre-assembled molecular aggregates may be used to construct network assemblies. The topologies of the resulting frameworks are directed by the geometry of the metals within the molecular SBUs and is also dependant on the ability of organic components to fill space effectively. These initial studies suggest that there is a rich chemistry surrounding the use of early main group metals in the supramolecular synthesis of extended frameworks.

## 6. ACKNOWLEDGEMENTS

We gratefully acknowledge the Petroleum Research Fund (41716-AC3) and the University of Notre Dame for support. We also thank the National Science Foundation for instrumentation support (CHE-0443233).

## 7. REFERENCES

[1] L. Brammer, *Chem. Soc. Rev.* **33** (2004) 476.
[2] D. Braga, *Chem. Commun.* (2003) 2751.
[3] S. L. James, *Chem. Soc. Rev.* **32** (2003) 276.
[4] C. Janiak, *Dalton Trans.* (2003) 2781.
[5] J. K. M. Sanders, *Chem. Eur. J.* **4** (1998) 1378.
[6] G. B. Gardner, D. Venkataraman, J. S. Moore and S. Lee, *Nature* **374** (1995) 792.
[7] S. J. Lee and W. Lin, *J. Am. Chem. Soc.* **124** (2002) 4554.
[8] H. O. Stumpf, L. Ouahab, Y. Pei, D. Grandjean and O. Kahn, *Science* **261 (**1993) 447.
[9] J. W. Steed and J. L. Atwood, *Supramolecular Chemistry*, (Wiley, New York, 2000).
[10] G. R. Desiraju, *Crystal Engineering. The Design of Organic Solids*, (Elsevier, Amsterdam, 1989).
[11] I. Haiduc and F. T. Edelmann, *Supramolecular Organometallic Chemistry*, (Wiley-VCH, New York, 1999).
[12] V. Chandrasekhar, S. Nagendran, S. Bansal, A. W. Cordes and A. Vij, *Organometallics* **21** (2002) 3297.
[13] O. M. Yaghi, M. O'Keeffe, N. W. Ockwig, H. K. Chae, M. Eddaoudi and J. Kim, *Nature* **423** (2003) 705.
[14] T. Stey and D. Stalke, D. in: *The Chemistry of Organolithium Compounds*, eds. Z. Rappoport and S. Patai (Wiley, New York, 2004), p 47.
[15] A. M. Sapse and P. v. R. Schleyer, *Lithium Chemistry, A Theoretical and Experimental Overview*, (Wiley, New York, 1995).
[16] E. Weiss, *Angew Chem. Int. Ed. Engl.* **32** (1993) 1501.
[17] An excellent special addition of the Journal of Solid-State Chemistry is available: O. M. Yaghi, M. O'Keefe, eds. *J. Solid State Chem.* **152** (2000) 1-321.
[18] K. W. Henderson, A. R. Kennedy, A. E. McKeown and D. S. Strachan, *J. Chem. Soc., Dalton Trans.* (2000) 4348.
[19] S. E. Denmark, K. A. Swiss and S. R. Wilson, *Angew. Chem. Int. Ed. Eng.* **35** (1996) 2515.
[20] A. F. Wells, *Three-Dimensional Nets and Polyhedra*, (Wiley-Interscience, New York, 1977).
[21] J. Kim, B. Chen, T. M. Reineke, H. Li, M. Eddaoudi, D. B. Moler, M. O'Keeffe and O. M. Yaghi, *J. Am. Chem. Soc*. **123 (**2001) 8239.
[22] G. Boche, *Angew. Chem. Int. Ed. Engl.* **28** (1989) 277.
[23] K. W. Henderson, A. R. Kennedy, L. Macdonald and D. J. MacDougall, *Inorg. Chem.* **42** (2003) 2839.
[24] K. W. Henderson, A. R. Kennedy and D. J. MacDougall, *Inorg. Chem.* **42** (2003) 2736.
[25] K. W. Henderson, A. R. Kennedy and D. J. MacDougall, *Phosphorous, Sulfur and Silicon* **179** (2004) 795.
[26] N. W. Ockwig, O. Delgado-Friedrichs, M. O'Keeffe and O. M. Yaghi, *Acc. Chem. Res.* **38** (2005) 176.
[27] D. Braga, F. Grepioni and G. R. Desiraju, *Chem. Rev.* **98** (1998) 1375.
[28] B. Walfort, L. Lameyer, W. Weiss, R. Herbst-Irmer, R. Bertermann, J. Rocha and D. Stalke, *Chem. Eur. J.* **7** (2001) 1417.
[29] R. E. Dinnebier, U. Behrens and O. Falk, *J. Am. Chem. Soc.* **120** (1998) 1430.
[30] For a specific example of 1,4-dioxane bridging in a lithiated coordination polymer see: O. Kühl, J. Sieler, G. Baum and E. M. Hey-Hawkins, *Z. Anorg. Allg. Chem.* **626** (2000) 605.
[31] T. J. Boyle, D. M. Pedrotty, T. M. Alam, S. C. Vick and M. A. Rodriguez, *Inorg. Chem*. **39** (2000) 5133.
[32] K. Thiele, H. Gorls and W. Seidel, *Z. Anorg. Allg. Chem*. **624** (1998) 1391.
[33] L. M. Jackman, D. Cizmeciyan, P. G. Williard and M. A. Nichols, *J. Am. Chem. Soc.* **115** (1993) 6262.
[34] P. K. Chattaraj, H. Lee and R. G. Parr, *J. Am. Chem. Soc.* **113** (1991) 1855.
[35] F. T. Deelmann, F. Pauer, M. Wedler and D. Stalke, *Inorg. Chem.* **31** (1992) 4143.
[36] M. Motevalli, D. Shah and A. C. Sullivan, *J. Chem. Soc, Dalton Trans*. (1993) 2849.
[37] R. Robson, in: *Comprehensive Supramolecular Chemistry*, eds. J. L. Atwood, J. E. D. Davies, D. D. MacNicol, F. Vögtle and K. S. Suslick, (Elsevier, Oxford, 1996).
[38] F. H. Allen, *Acta Crystallogr.* **B58 (**2002) 380.
[39] For example see: B. W. Skelton, C. R. Whitaker and A. H. White, *Aust. J. Chem.* **43** (1990) 755.

[40] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, C. G. Zakrzewski, J. A. Montgomery Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman [illegible]7[illegible]cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, [illegible] Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, J. L. Andres, M. Head-Gordon, E. S. Replogle and J. A. Pople, *Gaussian 98*, revision A.6; Gaussian, Inc.: Pittsburgh, PA, 1998.
[41] W. J. Hehre, R. Ditchfield and J. A. Pople, *J. Chem. Phys.* **56** (1972) 2257.
[42] P. C. Hariharan and J. A. Pople, *Theo. Chem. Acta* **28** (1973) 213.
[43] A. D. Becke, *J. Chem. Phys.* **98** (1993) 5648.
[44] A. D. McLean and G. S. Chandler, *J. Chem. Phys.* **72** (1980) 5639.
[45] L. M. Jackman and B. D. Smith, *J. Am. Chem. Soc.* **110** (1988) 3829.
[46] D. J. MacDougall, J. J. Morris, B. C. Noll and K. W. Henderson, *Chem. Commun.* (2004) 456.
[47] S. Mecozzi and J. Rebek Jr., *Chem. Eur. J.* **4** (1998) 1016.
[48] R. E. Dinnebier, M. Pink, J. Sieler and P. W. Stephens, *Inorg. Chem.* **36** (1997) 3398.
[49] T. J. Boyle, N. L. Andrews, M. A. Rodriguez, C. Campana and T. Yiu, *Inorg. Chem.* **42** (2003) 5357.
[50] M. Kunert, E. Dinjus, M. Nauck and J. Sieler, *Chem. Ber.* **130** (1997) 1461.
[51] D. J. MacDougall, B. C. Noll and K. W. Henderson, *Inorg. Chem.* **44** (2005) 1181.

# MODELING STUDIES OF ANTI-AMYLOID ANTIBODIES BOUND TO Aβ

Anna S. Gardberg, Nathan Tanner, Chris Dealwis*
*University of Tennessee at Knoxville, Department of Biochemistry and Molecular Biology, M407 Walters Life Sciences*
*Knoxville, Tennessee 37996-0830*

Israel Huff, Kimberly Salone, Ronald Wetzel
*University of Tennessee at Knoxville, Graduate School of Medicine, Knoxville, Tennessee, USA*
*Corresponding author: cdealwis@utk.edu
Keywords: A-beta, beta-amyloid, monoclonal antibodies, Alzheimer's, neurodegenerative disorders Autodock, protein-protein interaction
Grant sponsors: AG, NT, IH, RW, and CD are supported by NIH grant 5R01NS046356-03. KS was supported by a grant from the American Health Assistance Foundation.

## 1. ABSTRACT

WO1 and WO2 are conformation-specific monoclonal IgMs that bind the fibril state of the amyloid Aβ peptide (1-40), as well as amyloid fibrils of other disease-related proteins. Significantly, these antibodies (Abs) do not bind the soluble, monomeric state of Aβ (1-40) or the precursor form of other amyloids. The Abs have been sequenced and compared with sequences in the Kabat database; though some unusual charged residues were noted in WO1 and WO2, none are at the CDR. Three-dimensional models of the Fv fragments of WO1 and WO2 were generated with Web Antibody Modeling. A trigonal prism and a two-strand structural model of the Aβ amyloid core were compared by docking each with the Fv models of WO1 and WO2. The results predict binding of WO1 and WO2 to the Aβ trigonal prism at the protofilament face consisting of residues 23-27 and of WO2 to the Aβ two-strand at the end of the elongating fibril. No positive prediction for WO1 and the Aβ two-strand was made. Binding experiments with WO1 were determined to be sensitive to salt and pH conditions, suggesting the importance of electrostatic interactions for binding; this result is consistent with the proposed docking of WO1 to the Aβ trigonal prism.

## 2. INTRODUCTION

The aggregation of normally soluble proteins into insoluble, unbranched fibrils is the underlying pathology of a family of diseases known as the amyloidoses.[1] The hallmark event in amyloidogenesis is a change in the secondary and/or tertiary structure of a normal, soluble protein, rendering it prone to self-assembly into highly ordered para-crystalline arrays: fibrils. More than 20 proteins have been clinically identified as precursors of amyloid fibrils *in vivo*. These include the amyloid precursor protein (APP), Islet amyloid polypeptide (IAPP), α-synuclein, transthyretin (TTR), immunoglobulin light chain (LC), polyglutamine-repeats, and prion proteins, that are associated with diseases such as Alzheimer's,[2-4] type II diabetes,[5,6] Parkinson's disease,[7-9] familial polyneuropathy,[10] light chain associated (AL) amyloidosis,[11-13] Huntington's disease,[4,14] and the spongiform encephalopathies.[15] By understanding the three-dimensional structure of such fibrils, we might design therapeutic agents to target them. It is notoriously difficult to extract structural information directly from amyloid fibrils, which are insoluble and non-crystalline, so a great deal of study has gone into fiber diffraction studies,[16-18] mutation studies,[19] microscopy,[17,20,21] proteolysis,[22,23] ESR,[24,25] NMR,[24,26] SANS,[27] and deuterium exchange with mass spectroscopy[28-31] in an ongoing attempt to elucidate the molecular structure of the Aβ fibrils associated with Alzheimer's disease.

There is no consensus model for Aβ structure, but most models incorporate cross-beta secondary structure. The cross-beta structure (in which the fibril axis is perpendicular to the chain direction) of Aβ models is derived from the 4.75 Å meridional reflection in fiber diffraction studies.[18] A stacked, parallel beta-sheet arrangement is suggested by Burkoth, Benzinger, and others,[24,27,32] while a pair of concentric cylinders was proposed by Perutz, *et al*.[33] An anti-parallel or stacked-hairpin beta-sheet arrangement is suggested by Callaway and others.[34-37]

One recent model of the Aβ protofilament, proposed by Guo, *et al,*[38] features a trigonal prism of stacked, parallel Aβ(15-36) polypeptides; the model is shown in **Figure 1a**. This trigonal prism protofilament model is consistent with threading analysis[38] and proline-scanning mutagenesis,[19] and the proposed fibril model (of 6 protofilaments) appears to be consistent with electron microscopy. Recent data from hydrogen exchange NMR[39] and cysteine scanning mutagenesis[40] have called portions of this construction into question.

Another proposed configuration of the Aβ protofilament consists of two stacks of two parallel strands twisting around one another along the fibril axis; this model, shown in **Figure 1b**, is consistent with data that was available in 2002, especially that from solid-state NMR studies.[26] It is also in line with the recent crystal structure of the amyloid form of a seven-residue fragment from yeast protein Sup35[41], which supports a double-beta-sheet structure. Both the two-strand and trigonal prism Aβ protofilament constructs are consistent with specific bodies of experimental data, but are not completely determined by experimental data. Both models continue to evolve with the acquisition of new data.

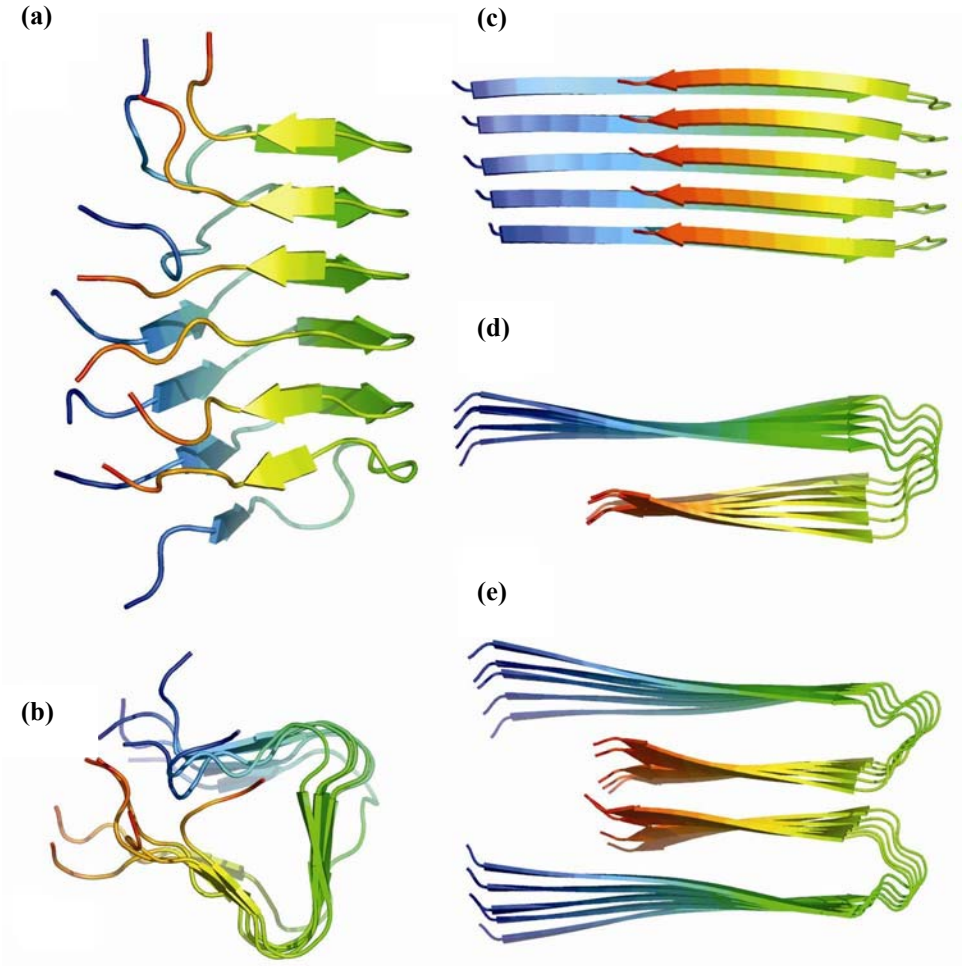



**Figure 1. (a)** Three-strand protofilament model in side view. **(b)** Three-strand, top view. **(c)** One half of proposed two-strand Aβ, side view. **(d)** Top view. **(e)** Two-fold symmetry about the *x*-axis has been used to generate one proposed configuration for the full protofilament from the two-strand model.

The monoclonal antibodies WO1 and WO2 (κ-light chain, murine, IgM) recognize a common conformational epitope shared by several different types of amyloid fibrils, with little dependence on amino acid sequence; the two bind the same antigen with similar affinities for the Aβ fibril.[42] This result opens the door to targeting whole classes of amyloid fibrils for detection and therapy.[43] Work to crystallize these antibodies (or their fragments) is ongoing. While IgM's are notoriously difficult to model,[44] the success of antibody modeling algorithms at predicting crystal structures of antibody variable regions[45] encouraged the use of computational models of WO1 and WO2 Fvs for docking simulations. Since WO1 and WO2 bind to Aβ amyloid, we hypothesized that a model of

the Aβ protofilament would, if correct, dock *in silico* to a model of the variable regions of WO1 and WO2. The configuration of the predicted complexes of the Fvs of antibodies WO1 and WO2 to two different models of Aβ protofilament is given. Experimental salt-dependence data is consistent with the prediction of WO1-Aβ binding for the trigonal prism model.

## 3. METHODS

### 3.1 Sequence Analysis

The nucleotide sequences of WO1 and WO2 were obtained by cloning and confirmed by multiple cloning using high fidelity polymerases, and further confirmed by extended N-terminal amino acid sequencing from analysis of the protein. Initial alignments were performed using Molecular Operating Environment (MOE™[46]), with Ab sequences obtained from the Protein Data Bank[47,48]. An extensive alignment was done using the Kabat sequence database testing program[49]. BLASTP analysis[50] was run on the WO1 and WO2 variable light chains using the Non-Redundant Protein Database at the San Diego Supercomputer Center. Canonical classes are based on Chothia nomenclature[51] and numbering scheme is that of Kabat[52]. The WO1 and WO2 Fv residue sequences were initially compared to one another. Then they were compared with sequences from the Kabat database to search for anomalies, unique features, or homologies to other antibodies. A sequence alignment of the CDRs is shown in Figure 2.

### 3.2 Model Building

Three-dimensional structures of WO1 and WO2 variable regions (Fv) were generated using the Web Antibody Modeling algorithm, WAM.[45] WAM is an improvement on the AbM program [53-56] with greater capability in modeling the highly variable H3 loop through a combination of knowledge-based and *ab initio* methods[45] . The WAM algorithm has a record of producing models that are 1.0-2.8 Å RMSD from observed structures for the heavy chain CDR3, and better than that for the canonical loops (typically 1.0-2.5 Å RMSD).[45]

### 3.3 Docking

#### 3.3.1 Preparation

A PDB file of the trigonal prism Aβ 1-40 hexameric protofilament model after molecular dynamics simulations ("Aβ trigonal prism", which models residues Gln15 through Val36)[38] was kindly provided by Juntao Guo and Ying Xu. A PDB file of the two-strand pentameric Aβ 1-40 protofilament model after energy minimization ("Aβ two-strand", which models residues 9 through 40)[26] was kindly provided by Robert Tycko.

The program Autodock3[57] predicts the interaction of ligands with macromolecular targets. As Autodock3 can handle no more than 2048 atoms, we reduced the two-strand protofilament to a tetramer and did not generate its symmetry pair. Control calculations performed with a protein-peptide complex, a protein-ligand complex, and a lysozyme-anti-lysozyme complex of known structure (PDB codes 1SMR, 1DQJ and 1DYI) showed that the charges assigned from forcefield calculations performed by AutoDockTools yielded less accurate results than those assigned by AMBER forcefield calculations[58,59] performed by InsightII.[60] We assigned charges using the following protocol. The atom names were standardized to Refmac5 conventions. Using Refmac5[61,62], hydrogen atoms were added in “riding” positions computed from the carbon, nitrogen, oxygen, and sulfur atomic positions. After capping the termini of each chain with (neutral) carboxylic and amine end groups, the partial charges for each atom (including polar hydrogen atoms) in each model were calculated with the InsightII program using the AMBER force field. Calculating the summed formal and partial charges for each model verified that the model is electrically neutral. A Sybyl-style file was then exported from InsightII.

#### 3.3.2 Computation

Autodock-style PDBQ files were prepared from the Sybyl-style files (described in the previous section) with *mol2topdbq* and *mol2topdbqs*, awk-based utilities packaged with Autodock3.[57] For purposes of solvation calculation, we designated the Fv models the “proteins” and the Aβ protofilament models the “ligands”. Despite the thermodynamic importance of side-chain motion,[63] neither the Fvs nor the Aβ models were permitted any torsional freedom; the Autodock package does not permit the side chains of the protein to move, and the Autotors utility for designating rotatable bonds in the ligand allows a maximum of 32, not nearly sufficient for a the 2395-2040-atom Aβ models. Grid parameter files were generated *via mkgpf3* (a script which comes packaged with

Autodock3) and adjusted to compensate for the large volume required for the interaction of the Fab's complementarity determining region (CDR) with the protofilament models. The grid maps were calculated using AutoGrid. The gridded area dimensions were 66 Å x 66 Å x 66 Å, with grid spacing increased to 0.55 Å. The final gridded regions encompassed the CDR of the Fv (as well as all of the heavy chain and most of the light chain of the model for WO1, and all save the C-termini for WO2) and the starting position of the Aβ model, with room for rotation and translation. Grid generation was performed with Autogrid3.
Docking was performed using the Lamarckian genetic algorithm (LGA), and the pseudo-Solis and Wets methods were applied for the local search. Each docking experiment was performed 50 times. Default parameters were applied for the docking experiments, with the exception of the initial rotation step size, which was set to 180. The results of the docking experiments were evaluated by calculating the positional root-mean-square (rms) deviation of the corresponding atoms of each conformation. Docking parameter files were prepared with the *mkdpf3* setup utility and adjusted by hand.

### 3.3.3 Contact Analysis

Contacts between the Fv models and the protofilament models in the predicted docking complexes were analyzed with the program CONTACT.[62] Docking clusters and configurations were examined Pymol,[64] with residue charge surfaces visualized. Electrostatic surface plots for publication were generated by the DelPhi module of InsightII.[60] Angles between solutions were computed by calculating the average direction vector along the protofilament axis for each solution and taking the inverse cosine of the dot-product of direction vectors.

### 3.4 Electrostatic Analysis

Electrostatic surfaces were calculated using DelPhi, a part of the InsightII graphical software package. The following parameters were employed: solvent dielectric 80, radius 1.4, ion radious 2.0, grid solvent extent 15, 65 points. Models were displayed and manipulated with InsightII and all computational work done on a Silicon Graphics computer system.

### 3.5 Salt Effects on Binding WO1

Binding of the WO1 antibody to Aβ fibrils was tested under 13 different conditions of cation, anion, salt concentration, and pH. 1xPBS buffer was used as a control. Experiments 1-5 comprised the salt-effects experiments; they consisted of: 5 mM HEPES at pH 7.5 with: 1) no salt, 2) 150 mM NaCl, 3) 600 mM NaCl, 4) 150 mM KCl, and 5) 75 mM $Na_2HPO_4$. In experiments 6-13, the effects of pH were examined; they consisted of:150 mM NaCl with 6) citric acid pH 3.0, 7) citric acid pH 5.8, 8) PIPES at pH 5.8, 9) PIPES at pH 7.4, 10) Bicine at pH 7.4, 11) Bicine at pH 8.5, 12) CHES at pH 8.5, and 13) CHES at pH 10.0.

No ionic interference was expected from HEPES, which has low ionic strength. $Na_2HPO_4$ has roughly double the ionic strength of the other two salts so it was used at half the concentration for a balanced comparison. The salt-effect experiments were buffered to pH 7.5, near physiological pH conditions. The pH effect experiments contained 150 mM NaCl to provide approximate physiological ionic strength conditions. The Aβ fibrils were prepared as described in *Kheterpal, et al, 2000*.[31]

## 4. RESULTS

### 4.1 Sequence Analysis

### 4.1.1 Comparison of WO1 to WO

The light chain variable regions share 66% identity (80% similarity) and the heavy chain variable regions share 82% identity (85% similarity). Overall charge composition of the Fv fragments is very similar, with isoelectric points (pIs) of the variable regions calculated to be: WO1 $V_H$ 7.3, $V_L$ 8.7, WO2 $V_H$ 7.2, and $V_L$ 8.7.

While both sequences are rich in hydrogen bonding residues, WO1 $V_L$ CDR has a greater proportion of charged Asp, Glu, Arg, and Lys residues (19% vs. 11%), while the WO2 $V_L$ CDR is richer in uncharged, polar hydroxyl residues (42% for WO1 $V_L$, 64% for WO2 $V_L$). The sequences of the CDR's are compared in **Figure 2**. The WO1 and WO2 heavy chain variable regions are more similar in overall sequence, and the number of charged residues in the $V_H$ is identical. Still, the two charged residues of WO1's $V_H$ are at its CDR, and those of WO2's $V_H$ are not at its CDR. WO1's light chain CDR contains charged residues at positions occupied by hydroxyl residues in WO2 such as Arg24, Lys52, and Asp56 in WO1 which are replaced by Thr, Ser, and Ser in WO2. Similarly, the unique charged

residue Arg93 in WO2 replaces a serine residue in WO1. WO1's greater proportion of charged residues and WO2's greater proportion of uncharged polar residues suggest that WO1 and WO2 bind amyloid differently from one another.

a)

| VL CDR1 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 31a | 32 | 33 | 34 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| WO1 | R | A | S | G | N | I | H | N | - | ***Y*** | L | A |
| WO2 | T | A | S | S | S | V | S | S | S | ***Y*** | L | H |
| α-Lyso | R | A | S | Q | N | I | S,H | N | N | - | L | H,A |
| Germ. | R,K | S,A | S | Q | S,N | I | nh* | nh | S | nh | L | nh |
| Kappa | R,K | A,T | S | Q | nh | V,I | nh | S,T | - | nh | L | nh |

| CDR2,3 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| WO1 | N | A | K | T | L | ***A*** | D | Q | H | F | W | S | T | P | ***Y*** | T |
| WO2 | S | T | S | N | L | ***A*** | S | L | Q | Y | H | R | S | P | ***Y*** | T |
| α-Lyso | Y | T | T,S | Q,T | S,L | nh | D,S | Q | Q | F,Y | nh | S | - | P | R,L | T |
| Germ | nh | A,V | S,A | nh | R,L | D,E | S | nh | Q | nh | S | nh | nh | P | R | T |
| Kappa | nh | A | S | nh | R,L | nh | S | Q | Q | S,Y | nh | S,D | nh | P | R,L | T |

b)

| VH CDR1,2 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| WO1 | G | Y | T | F | T | E | Y | ***T*** | M | H | G | I | ***N*** | P | N | ***N*** | G | T | **S** |
| WO2 | G | Y | S | F | T | G | Y | ***T*** | M | N | L | I | ***N*** | P | Y | ***N*** | G | T | **S** |
| α-Lyso | G | Y | T,S | F | T | T | Y | W | I,V | E,S | E | I | L | P | G | S | nh | T | Y,D |
| Germ | G | F | T | F | T | D | Y | Y | M | S,H | F | I | R | nh | nh | nh | nh | T | E,K |
| Kappa | G | F,Y | T | F | T | D | Y | Y | I,M | H,K | nh | I | nh | P | nh | nh | G,S | T | nh |

| CDR3 | 96 | 97 | 98 | 99 | 100 | 101 | 102 | 103 | 104 | 105 | 106 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| WO1 | C | A | R | D | D | G | - | ***Y*** | Y | G | F |
| WO2 | C | A | R | R | L | G | D | ***Y*** | Y | A | M |
| α-Lys | C | A | R,S | nh | D,E | G | - | nh | Y | G | nh |
| Germl | C | A | R,S | D | nh | nh | nh | nh | Y | A,Y | M,G |
| Kappa | C | A | nh | G,D | nh | G | nh | nh | nh | nh | F |

**Figure 2.** CDR composition of a) $V_L$s and b) $V_H$s compared with those of anti-lysozyme, germline, and kappa-light-chain antibodies. **nh* signifies no homology. Bold, italicized text highlights conserved residues in WO1 and WO2 White-on-black text indicates identical or homologous residues for WO1, WO2, and one or more compared sequences

### 4.1.2 Light Chain

A preliminary sequence alignment was performed using a database of various antibody sequences, including subsets of anti-lysozyme Ab chains and germline Ab chains. The results showed a high degree of homology between heavy and light chain sequences, with several residues noted as distinctive in **Table 1**. The CDRs of the light and heavy chains of WO1 and WO2 are rich in hydrogen-bonding residues relative to anti-lysozyme and germline antibodies, suggesting that hydrogen bonding plays an important role in amyloid-recognition.

A more extensive alignment was done using the Kabat sequence database testing program. The WO1 $V_H$ and $V_L$ sequences were compared to the database of 2707 light chains and 3471 heavy chains. The unique residues Lys42 and Lys74 had no consensus within the database; at these positions, which are not at the CDR, Lys occurred in less than 1% of the sequences. Lys52 at light chain CDR2, however, was a hydroxyl residue in the majority of the 2701 chains. None of the WO2-unique charged or hydroxyl residues showed any incongruity.

| Light Chain Residue WO1 WO2 | Equivalents | Residue exhibited by chains with non-identity |
|---|---|---|
| Lys42 Ser43 | No | None charged, only Gly, Gln, Tyr |
| Lys52 Ser53 | Lys only in WO1 | No charged, 95% Tyr or Ser |
| Asp56 Ser57 | Asp in WO1 and (AL) | 99% Pro, (G) have Thr/Ser |
| Lys74 Thr75 | Lys only inWO1 | 100% Glu/Asp, (G) has Thr |

**Table 1:** Light chain-unique residues and comparison for WO1 residues that are less than 10% homologous to database sequences in the initial alignment tests. The alignment was done with specific interest in the anti-lysozyme Abs as well as germline Abs, and any residues here are noted if appearing in either chain. The residue is listed with its location, as well as the residues that were most commonly shared throughout the other Ab sequences. The equivalent residue from WO2 is also included for comparison. Codes: antilysozyme (AL), germline (G).

### 4.1.3 Heavy Chain

Through the preliminary alignment including the anti-lysozyme, germline, and kappa-light chain sequences, several residues were noted as distinctive. They are listed in **Table 2**. As with the light chain results, the most striking disparities between the WO1/WO2 heavy chains and comparison germline and anti-lysozyme sequences occur with charged residues. Again, the Kabat alignment was performed. The equivalent positions for residues Lys63, Lys65, Lys67, Lys74, Asp73 (in WO1 and WO2) and Arg84 (in WO1 only) have no consensus, but these charged residues are not unique and do not occupy the CDR. The only significant discrepancy was the Asn44 of the WO2 heavy chain. The Asn residue found at this site is found in only 13 examples, 0.326% of the database. Typically, a Ser residue is at this position. This position is also not at the CDR.

| Heavy Chain Residue WO1 WO2 | Equivalents | Residue exhibited by chains with non-identity |
|---|---|---|
| Lys63 Lys63 | WO1/WO2, (G) | No other charged residues |
| Lys65 Lys65 | WO1/WO2, (G), (AL) | 100% Val, Leu, or Phe |
| Lys67 Lys67 | WO1/WO2, (G), (AL) | 100% Gly, Ser, Asp |
| Asp73 Asp73 | WO1/WO2, (G), (AL) | Mostly Asp/Glu |
| Lys74 Lys74 | WO1/WO2, (G) | 99% Asp/Glu |
| Arg84 Leu84 | Arg only in WO1 | Non-consensus, but no other charged |

**Table 2:** Heavy chain unique residues and comparison. The residues displaying less than 10% homology in the initial alignments to database sequences are included here. Any germline or anti-lysozyme Ab commonalities are noted, as well as the residues which occur in the chains which are not homologous to the WO1/WO2 heavy chains. Codes: antilysozyme (AL), germline (G).

### 4.1.4 Fv Models

Of the WO1 light chain's unique lysine residues (42, 52, and 74), only Lys52 is at the CDR and may play a role in docking. However WO1 light-chain Asp56 (unique to WO1 and anti-lysozyme Abs) may play a role, since it makes a hydrogen bond to an Asn residue in the trigonal prism model.

### 4.2 Electrostatic Surface Potentials for WO1 and WO2

WO1 and WO2 Fvs display regions of significant charge density (as calculated by the DelPhi module of InsightII), observed in the electrostatic potential map shown in Figure 3. Though the pIs of the WO1 and WO2 Fvs are nearly neutral, they clearly have highly charged regions. The presence and distribution of these charged segments indicates a possible role of charge interactions in antigen binding. The line of positively charged residues on each of the light-chains suggests a template for hydrogen bonding and salt bridges. The line of negatively charged residues at the heavy-light interface of WO1's CDR (as seen in **Figure 3**) reflects the higher proportion of negatively charged residues in its composition.

### 4.3 Electrostatic Surface Potentials for the Protofilament Models

The most striking features of both models are the bands of charged residues aligned along the stacking axis; these representations are shown in **Figure 4**. The charge bands are a consequence of the in-register parallel-stacking of the models. Labeling the trigonal prism protofilament model's face containing residues 17-22 "A", 23-27 "B", and 28-36 "C" (Aβ 1-40 numbering basis), we see that faces A and C display exposed hydrophobic residues. In the two-strand model, the N-terminal face consists of residues 9-23; this corresponds to the "A" face of the trigonal prism

model. There is then a 180º turn consisting of residues 24-27; this turn (instead of a β sheet) is in sharp contrast to the "B" face of the other model. Finally, residues 30-40 form the C-terminal face of the two-strand model.

**(a)** **(b)**

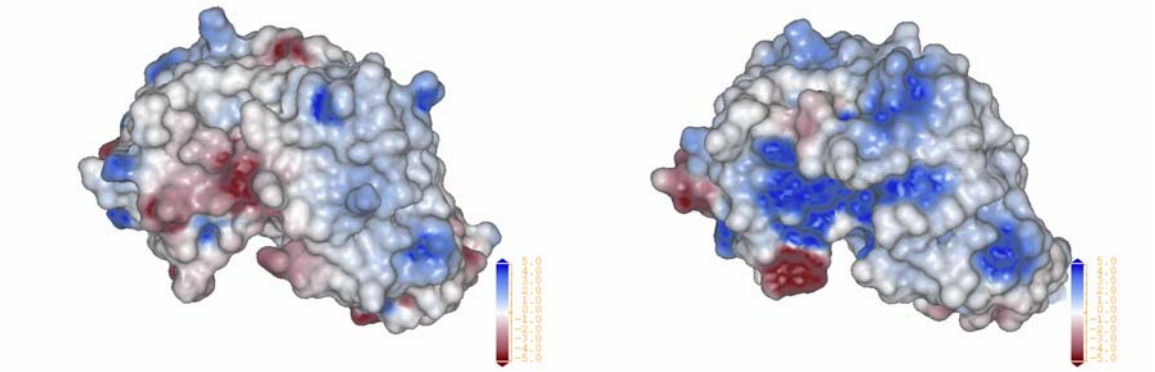


**Figure 3**. Electrostatic surfaces for WO1 **(a)** and WO2 **(b)** Fv models. The viewer faces the CDRs. In each image, the heavy chain fragment is on the left.

**(a)**

**(b)**

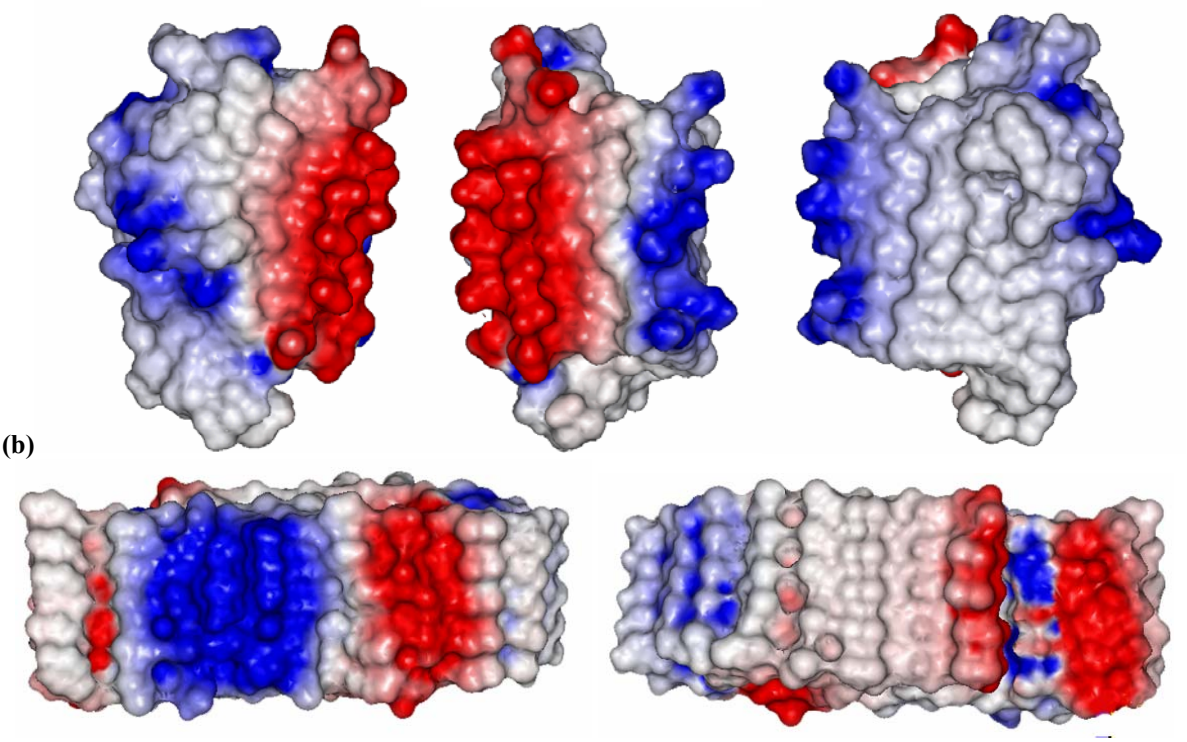


**Figure 4**. **(a)** Electrostatic surfaces for the A, B, and C faces of the trigonal prism protofilament model. The line of Glu22 and Asp23 residues (shown in red) at the corner of the A and B faces is particularly striking. **(b)** Electrostatic surfaces for the two faces of the two-strand model. The N-terminal face is shown in the left-hand image; strand direction is left to right. The C-terminal face is shown in the right-hand image; strand direction is left to right.

### 4.4 Docking WO1-Trigonal Prism-Model

AutoDock3 predicted 50 binding configurations. Both the Fv and the trigonal prism model were treated as rigid bodies. Autodock3 found 41 distinct conformational clusters (using an RMSD-tolerance of 1.0 Å), of which 8 contained more than one solution. The variation in energy from the best to the worst solution was 7.0 kcal/mol.

Although there was one cluster of two solutions in the top 7 (with the B face laterally at the heavy-light interface), more intriguing is a collection of related clusters beginning at the 8th-ranked solution. In total, 21 solutions have the B face of the trigonal prism protofilament model longitudinally oriented at the light-chain CDR, with some overlap onto heavy chain CDR. Some are offset by one, two, or three monomers (i.e., the stack of six Aβ monomers is shifted along the stacking axis), others by angles less than 27˚. This is the largest supercluster; its lowest-energy solution is presented in **Figure 5a**. The particular solution used for the following analysis was the lowest-energy member of this supercluster, which is 3.5 kcal/mol (~1 hydrogen bond) less favorable than the lowest-energy solution of all docking runs.

### 4.4.1 Salt Bridges and Ion Pairs

This docking model indicates that the line of Glu22 and Asp23 residues in the trigonal prism model binds to the corresponding line of polar and positively charged residues of WO1's light-chain CDR through salt bridges and type 2 hydrogen bonds between acidic residues and carbonyl groups. The line of Lys28 residues likewise interact with the trail of polar and negatively charged residues of the light *and* heavy chain CDRs. Breaking the ion pairs down by Szilyagi and Zavodsky classification,[65] there are four bonds at < 4.0 Å, six bonds between 4.0 and 6.0 Å, and ten pairs between 6.0 and 8.0 Å. *Hydrogen bonding*. The O and N atoms of the side chains of the line of Asn27 residues on the three-strand model make hydrogen bonds with Tyr and Asp residues of WO1. The carbonyl O atoms of the line of (inward-pointing) Val24 and Gly25 residues bond to Tyr residues of WO1. The backbone N of inward-pointing Ser26 makes some possible hydrogen-bonds. There are 17 likely hydrogen bonds at less than 3.5 Å, two of which were also counted as close ion pairs. The hydrogen bonding interface is shown in **Figure 5d**.

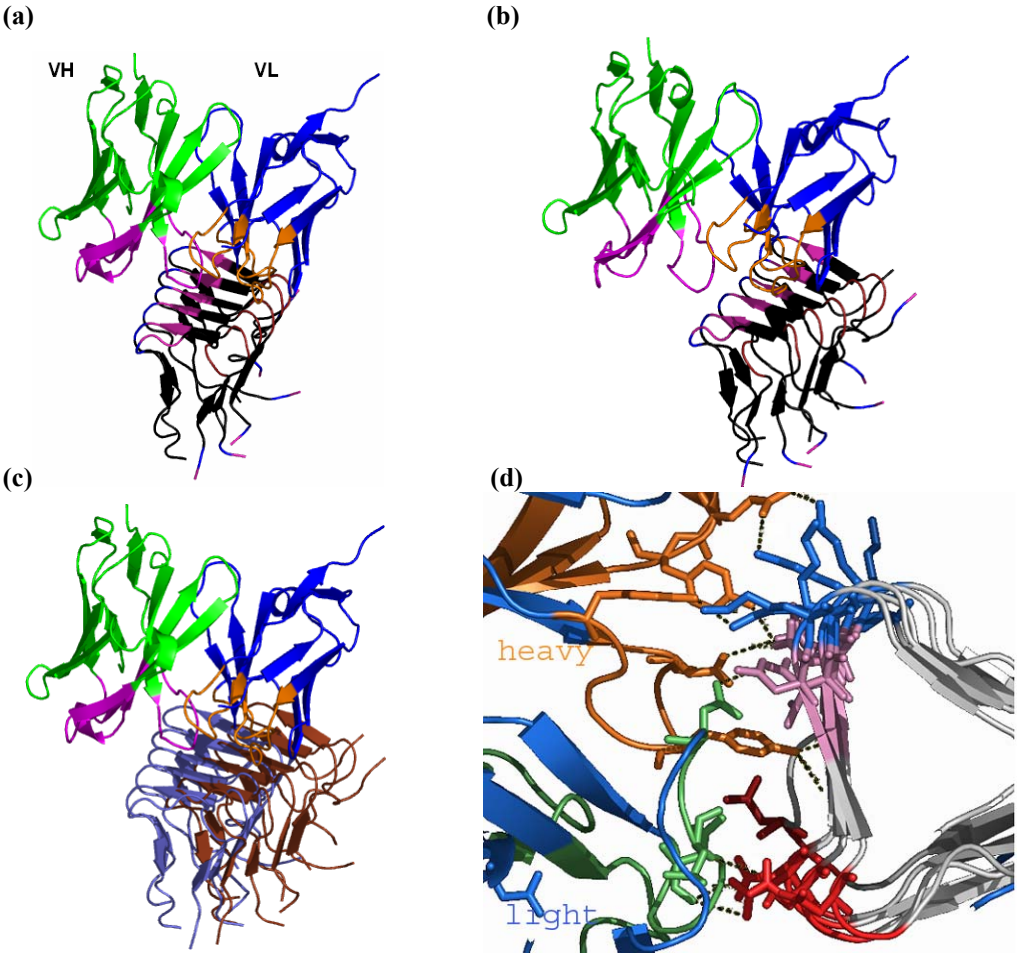



**Figure 5**. **(a)** WO1 and **(b)** WO2 Fv docked with trigonal prism protofilament model in the predicted configuration (top solution in supercluster). The heavy chain is on the left, the light chain is on the right, and the CDRs are highlighted. The trigonal prism model is shown colored according to residue: negative, red; positive, blue; polar, pink; hydrophobic, black. **(c)** The two solutions are showed overlain. **(d)** A close-up view of the docking interface for the WO1-trigonal-prism model. The color code is as for **(a)** and **(b)**, but hydrophobic residues are shown in grey and hydrogen bonds are shown in black.

**4.4.2 Hydrophobic Interactions**
There are no aromatic rings on the B-face of the trigonal prism model, and the hydrophobic Val24 side chain points toward the interior of the trigonal prism model. Thus hydrophobic interactions between WO1 and the trigonal prism model are not observed in this docking model.

**4.5 Docking WO2-Trigonal Prism-Model**
AutoDock3 predicted 50 binding solutions. The Fv and the trigonal prism model were both treated as rigid bodies. For WO2 and the trigonal prism protofilament model, Autodock3 found 42 distinct conformational clusters (using an RMSD-tolerance of 1.0 Å), of which 5 contained more than one member. The variation in energy from the best to the worst solution was 9.3 kcal/mol. There was one cluster of two solutions in the top 13, but the largest cluster began with the $14^{th}$-ranked solution. Within 1.0 Å R.M.S.D., there are 5 similar solutions. In total, 11 solutions have the B face of the trigonal protofilament model at the light-chain CDR, with slight overlap onto heavy chain CDR. One of them is offset by one monomer, others by acute angles. This is the largest supercluster; its lowest-energy solution is shown in **Figure 5b**. This solution resembles the supercluster solution found in the WO1 docking, but makes a 16° angle to it. The particular solution used for the following analysis was the lowest-energy member of this supercluster, which is 6.4 kcal/mol less favorable than the lowest-energy solution.

**4.5.1 Salt Bridges and Ion Pairs**
The analysis for this docked conformation is very similar to that of the WO1-Aβ6 solution, but there are no salt bridges at less than 4.0 Å, only two ion pairs between 4.0 and 6.0 Å, and five between 6.0 and 8.0 Å. *Hydrogen Bonding and hydrophobic interactions.* Thirteen likely hydrogen-bonds at less than 3.5 Å exist between WO2's Fv model and the docked trigonal-prism model. No hydrophobic interactions were observed for this docking model.

**4.6 Docking. WO1-Two-Strand-Model**
AutoDock3 predicted 50 binding solutions. Both the Fv and the protofilament model were treated as rigid bodies. For WO1 and the two-strand protofilament model, Autodock3 found 4 distinct conformational clusters with more than two members (using an RMSD-tolerance of 2.0 Å. The variation in energy from the best to the worst solution was 5.6 kcal/mol. In total, 4 solutions have the N-terminal face on the CDR, 4 have the open end of the two-strand model sitting longitudinally on the CDR, 7 solutions have the "bottom" end (right-hand-rule) of the two-strand protofilament model laterally across the CDR, and 3 have the closed end of the two-strand model near the CDR. The several clusters of near-equal energy suggest a negative result; they are shown in **Figure 6a**. Two of these clusters are not predicted to make any salt bridges at < 4 Å, but the other two are expected to make one and two such bonds, respectively.

**4.7 Docking. WO2-Two-Strand-Model**
As before, AutoDock3 predicted 50 binding solutions with both components (the Fv and the protofilament model) treated as rigid bodies. For WO2 and the two-strand protofilament model, Autodock3 found 35 distinct conformational clusters (using an RMSD-tolerance of 2.0 Å), of which 3 contained more than two members. The variation in energy from the best to the worst solution was 8.6 kcal/mol. The top 7 solutions have the "bottom" end (right-hand-rule) of the two-strand protofilament model laterally across the CDR. This cluster averages 3.8 kcal/mol lower in energy than the next best solution and it is the largest cluster, indicating a positive result; it is displayed in **Figure 6b**. This somewhat resembles the 7-member cluster for WO1-two-strand-model (shown in **Figure 6a**). The particular solution used for the following analysis was the lowest-energy member of the cluster.

**4.7.1 Salt Bridges, Ion Pairs, Hydrogen Bonding, and Hydrophobic Interactions**
No salt bridges were observed out to 4.0 Å, but 10 ion pairs in the range of 4-6 Å suggest a modest electrostatic contribution to WO2 binding Aβ. WO2 and the two-strand model make 7 likely hydrogen bonds at less than 3.4 Å. No *pi*-overlaps are observed.

**4.8 Docking Controls**
Positive controls #1 and #2 used rigid models, positive control #3 allowed flexibility in the ligand. Positive control #1 successfully recreated the crystal structure 1DQJ,[66] lysozyme complexed to antilysozyme. Two of the 50 solutions clustered near that from the crystal structure. These two averaged 11.8 kcal/mol lower in energy than the next most energetically favorable solution. Positive control #2, mouse renin complexed to the angiotensin analog CH-66 (PDB code 1SMR),[67] showed a similar result, with the 44-member dominant cluster 19.7 kcal/mol lower in energy. Positive control #3 did not precisely recreate the crystal structure 1DYI,[68] DHFR complexed with

folate, but rather predicted 10 similar solutions for the 32-atom folate molecule in the correct binding pocket within 2.3 to 10.8 Å RMSD of the true crystal structure, and within 3.5 kcal/mol of one another. Negative control #1 consisted of "docking" the trigonal prism protofilament model with the crystal structure of mouse renin. The result consisted of three superclusters (23, 14, and 5 members) of solutions whose lowest-energy members were within 5.0 kcal/mol of each other. Negative control #2 consisted of "docking" the WO1 Fv model with the inhibitor from the mouse renin crystal structure. All fifty top solutions were within 4 kcal/mol of one another. There was one two-member cluster and one four-member cluster. Solutions ranged all over the Fv.

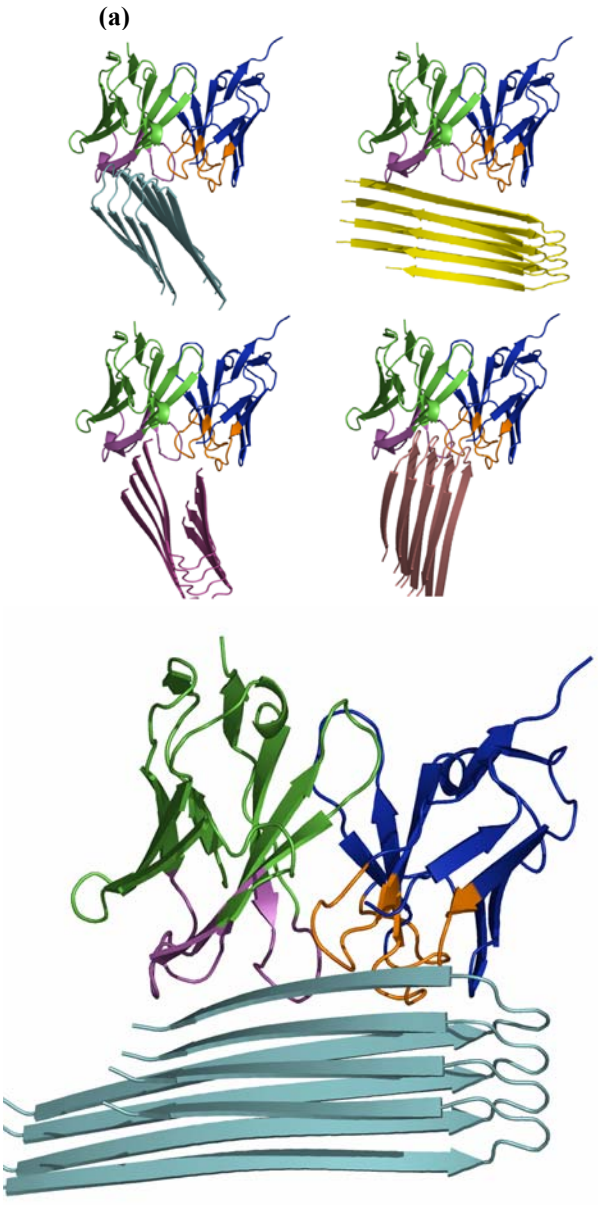



**Figure 6.** **(a)** WO1 Fv docked with two-strand protofilament model in the four predicted configurations. **(b)** WO2 Fv docked with two-strand protofilament model in the top predicted configuration

### 4.9 Binding vs. Salt Concentration

Binding constants were calculated from the sigmoid midpoints of the graph in **Figure 7**. The strongest binding of WO1 and Aβ was observed in PBS; the $EC_{50}$ was 2 nM. Of the 5 mM Hepes conditions, the ones at 150 mM NaCl (nearly-physiological) and at no salt resulted in the strongest binding of WO1 and Aβ, 4 nM. In the presence of 600

mM NaCl, the $EC_{50}$ constant changed sharply to 20 nM, indicating five-fold less binding. Moreover, the 150 mM KCl and 75 mM $Na_xH_yPO_4$ conditions resulted in 8 nM and 10 nM EC50's, respectively. The significantly reduced binding in high salt suggests that electrostatic interactions are relatively important to binding and hydrophobic interactions are relatively unimportant.

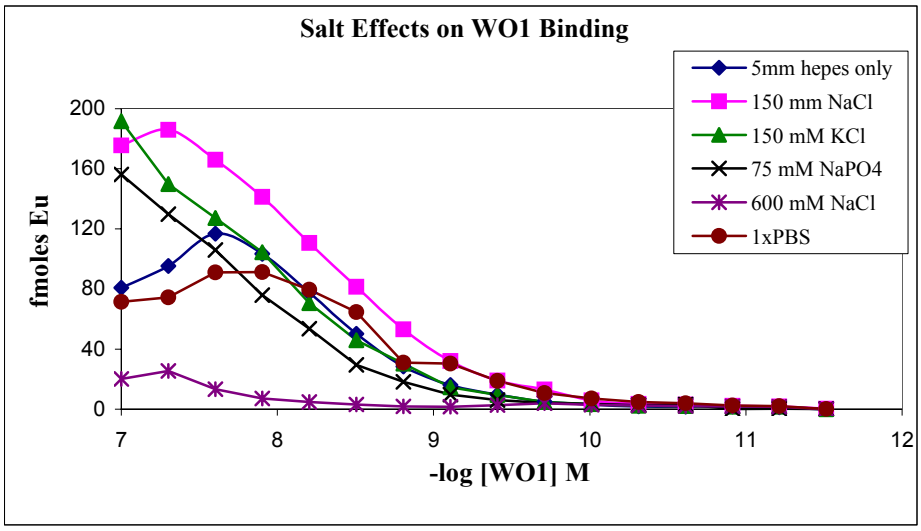



**Figure 7. Salt effect on WO1 binding.** Binding under various salt conditions was calculated from sigmoid midpoints. The $EC_{50}$ values are given in the text.

WO1 binding to Aβ was tested in a range of pH conditions. The Citrate pH 3.0 condition resulted in essentially no binding. Citrate and PIPES at pH 5.8 showed $EC_{50}$ values of 2 nM and >20 nM, respectively. The $EC_{50}$ values for PIPES and Bicine at pH 7.4 were 2 nM and 1.5 nM. For Bicine and CHES at pH 8.5, $EC_{50}$ values were 16 nM and 10 nM. For CHES at pH 10.0, the $EC_{50}$ was > 50 nM. (**Figure 8**). The data suggest that WO1 binds Aβ amyloid best closer to physiological pH values. If hydrophobic interactions dominated WO1- Aβ amyloid binding, this would likely not be the case. This supports the salt effect results suggesting that binding depends at least in part on electrostatic interactions.

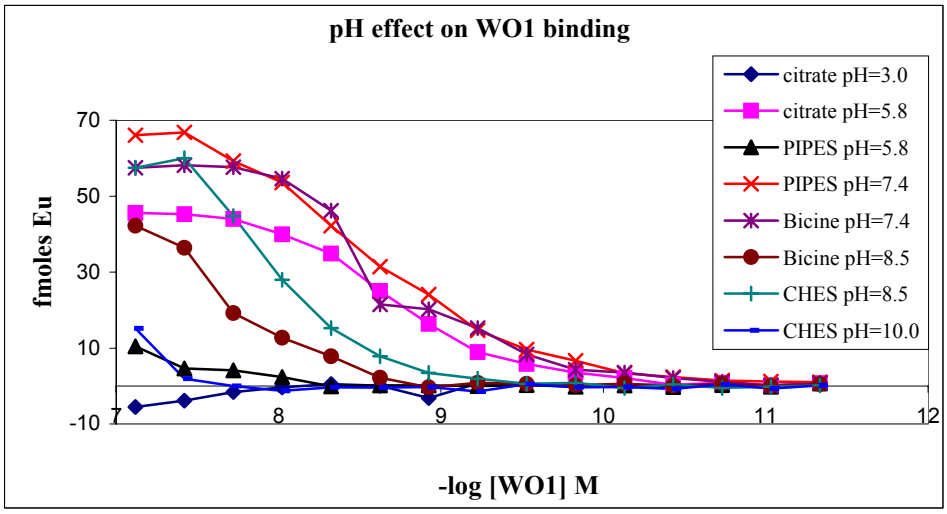



**Figure 8. pH effect on WO1 binding.** The show binding under various pH conditions. Binding affinities were calculated from the sigmoid midpoints. The $EC_{50}$ values are given in the text.

## 5. DISCUSSION

### 5.1 Comparison with Experimental Results

The model of WO1- trigonal prism-model docking that we have proposed, shown in **Figure 5a**, features many hydrogen-bonding pairs and salt-bridges; that agrees with the experimental salt-effect data for WO1- Aβ binding. However, the salt-binding data itself is at odds with WO1's general amyloid-recognition behavior which includes binding to (uncharged) polyglutamine fibrils. This prediction is also called into question by recent data from hydrogen exchange NMR[39] and cysteine scanning mutagenesis[40]. These data are not straightforward to interpret, but they suggest that residues 23-27 (the so-called "B face" of the trigonal-prism protofilament model) may not be in ordered β-sheet structure. There are no strong salt bridges at less than 4.0 Å predicted for binding of WO2 to either the trigonal prism model or the two-strand model.

### 5.2 Assemblies

The Fv models presented here are docked only to a protofilament not to the full Aβ fibril, but these docked models are consistent with the hierarchical fibril model for Aβ amyloid proposed by their authors.[26,38] The fibril model based on the trigonal prism protofilament construction consists of six trigonal prisms packed vertically against one another in two rows of three, in which each protofilament face is exposed twice on the outside surface of the fibril model. For the two-strand model, a full protofilament is generated by applying a two-fold symmetry axis along the C-terminal face perpendicular to the stacking axis of the model. In this way, the open ends of both stacks of monomers point the same way as the stacks twist helically around one another. One depiction of this twisting stack may be found in this work, **Figure 1e**; another is in Petkova, *et al* PNAS 2002, figure 5a. This implies that the C-terminal face of the model is not available for binding; as the predicted docking mode does not employ the C-terminal face, this does not affect our analysis. The Fv model likewise has a single antigen-recognition region for WO1/WO2, whereas the physiological IgM presents ten recognition sites, resulting in high avidity and increased total binding strength.

### 5.3 Implications of Trigonal Prism Results

The docking results for the trigonal prism model suggest that WO1 and WO2 bind to the face of amyloid fibrils similarly. The in-register parallel beta-stacking of the trigonal prism model requires that each residue of the polypeptide line up with its equivalent in the next layer of the prism. Thus, while Glu22 makes peptide bonds to Ala21 and Asp23, it makes hydrogen bonds to Glu22 on neighboring strands. Viewing a face of the prism, then, one would see lines of equivalent residues. The line of Glu22 and Asp23 residues (**Figure 4**) is predicted to bind to the corresponding line of positive and polar residues of WO1's light-chain CDR. The line of Asn27 and Lys28 residues likewise interact with the trail of negative and polar residues.

The lineup of positive and negative charges on the B face with those at the CDR suggests another available binding mode. The A face of the model also has such a lineup of charges (Figure 3). The N-terminal residues of the trigonal prism model occlude part of the A face in this model, which was not allowed torsional freedom, so the A face was less accessible for binding than the B face. Even so, 2/50 solutions for WO1 and 5/50 solutions for WO2, though not clustered, did predict some form of docking to the A face. It is worth noting, furthermore, that earlier simulations with an un-minimized model (in which the N terminal residues took other orientations) yielded a supercluster of docking solutions at the A face of the trigonal prism model with WO1. It is thus likely that a more computationally intensive docking simulation, one which allowed rotation around the bonds of the trigonal prism model, would have shown a bimodal distribution of solutions, one at the A face and another at the B face. These facts, coupled with the new data that suggest non β-sheet conformation for the B-face,[39,40] to some degree weaken the case for the trigonal-prism model for Aβ protofilament. Therefore WO1/WO2 binding to the A face of Aβ remains a distinct possibility. As only 4/50 solutions featured the C face, and they all took very different orientations, our current docking model suggests that the C face is unlikely to feature in WO1 binding.

### 5.4 Implications of Two-Strand Results

Docking calculations based on the two-strand model do not indicate a conclusive result for WO1-binding. However, computations for WO2 support WO2 binding to the end of elongating amyloid fibrils, while only the 48$^{th}$ ranked hit for WO2 and the trigonal prism resembles such an arrangement. This contrasts with expectation that the lines of charge along the N-terminal face would provide a recognition site for the antibodies, but is consistent with the fact that these antibodies bind amyloid fibrils (such as polyglutamine) that lack charged residues. Also, the lower degree

of order found in residues 23-27 in the two-strand model may be more consistent with the new hydrogen exchange and cysteine mutagenesis data.

## 6. CONCLUSIONS

The sequences of two amyloid-recognizing IgMs were compared and analyzed. Structural models of their Fvs were constructed. WO1's Fv was found to be unusually rich in charged residues, while WO2 is rich in Ser residues. As one might expect from the differences in their secondary and ternary structures, the two different protofilament models lead to different predicted docking models. Results of docking simulation for the trigonal-prism model imply that WO1 and WO2 bind to the face of a fibril of Aβ. The results for the two-strand model make no positive prediction for WO1, but results for the two-strand model suggest that WO2 binds to the ends of fibrils. The results for WO1 and the trigonal prism model agree with experimental salt-binding data, while the results for WO2 and the two-strand model are more consistent with WO1/WO2's diverse binding behavior. While these results do not support either model over the other, they make testable predictions.

## 7. ACKNOWLEDGMENTS

We thank Doctors Juntao Guo and Ying Xu for the trigonal prism protofilament model and Doctor Robert Tycko for the two-strand protofilament model.

## 8. REFERENCES

[1] R. H. Falk, R. L. Comenzo and M. Skinner, *N Engl J Med* 337 (1997).
[2] D. J. Selkoe, *Ann N Y Acad Sci* 924 (2000), 17.
[3] I. Tariska, *Acta Med Acad Sci Hung* 21 (1965), 483.
[4] P. A. Temussi, L. Masino and A. Pastore, *Embo J* 22 (2003), 355.
[5] E. T. Jaikaran and A. Clark, *Biochim Biophys Acta* 1537 (2001), 179.
[6] M. R. Hayden and S. C. Tyagi, *Jop* 2 (2001), 124.
[7] O. Corti and A. Brice, *J Soc Biol* 196 (2002), 95.
[8] R. Kruger, O. Eberhardt, O. Riess and J. B. Schulz, *Trends Mol Med* 8 (2002), 236.
[9] J. M. George, *Genome Biol* 3 (2002), REVIEWS3002.
[10] M. Pras, F. Prelli, E. C. Franklin and B. Frangione, *Proc Natl Acad Sci U S A* 80 (1983), 539.
[11] H. D. Feiner, *Hum Pathol* 19 (1988), 1255.
[12] J. N. Buxbaum, J. V. Chuba, G. C. Hellman, A. Solomon and G. R. Gallo, *Ann Intern Med* 112 (1990), 455.
[13] V. Bellotti, P. Mangione and G. Merlini, *J Struct Biol* 130 (2000), 280.
[14] C. A. Ross, *Neuron* 35 (2002), 819.
[15] C. Weissmann, *Nat Rev Microbiol* 2 (2004), 861.
[16] H. Inouye, P. E. Fraser and D. A. Kirschner, *Biophys J* 64 (1993), 502.
[17] M. Sunde and C. Blake, *Adv Protein Chem* 50 (1997), 123.
[18] S. B. Malinchik, H. Inouye, K. E. Szumowski and D. A. Kirschner, *Biophys J* 74 (1998), 537.
[19] A. D. Williams, et al., *J Mol Biol* 335 (2004), 833.
[20] J. D. Harper, S. S. Wong, C. M. Lieber and P. T. Lansbury, Jr., *Biochemistry* 38 (1999), 8972.
[21] L. C. Serpell, et al., *J Mol Biol* 300 (2000), 1033.
[22] T. C. Saido, W. Yamao-Harigaya, T. Iwatsubo and S. Kawashima, *Neurosci Lett* 215 (1996), 173.
[23] I. Kheterpal, A. Williams, C. Murphy, B. Bledsoe and R. Wetzel, *Biochemistry* 40 (2001), 11757.
[24] O. N. Antzutkin, R. D. Leapman, J. J. Balbach and R. Tycko, *Biochemistry* 41 (2002), 15436.
[25] M. Torok, et al., *J Biol Chem* 277 (2002), 40810.
[26] A. T. Petkova, et al., *Proc Natl Acad Sci U S A* 99 (2002), 16742.
[27] T. S. Burkoth, et al., *J. Am. Chem. Soc.* 122 (2000), 7883.
[28] I. Kheterpal, et al., *Biochemistry* 42 (2003), 14092.
[29] S. S. Wang, S. A. Tobler, T. A. Good and E. J. Fernandez, *Biochemistry* 42 (2003), 9507.
[30] M. Kraus, M. Bienert and E. Krause, *Rapid Commun Mass Spectrom* 17 (2003), 222.
[31] I. Kheterpal, S. Zhou, K. D. Cook and R. Wetzel, *Proc Natl Acad Sci U S A* 97 (2000), 13597.
[32] T. L. Benzinger, et al., *Proc Natl Acad Sci U S A* 95 (1998), 13407.
[33] M. F. Perutz, J. T. Finch, J. Berriman and A. Lesk, *Proc Natl Acad Sci U S A* 99 (2002), 5591.
[34] L. O. Tjernberg, et al., *Biochem J* 366 (2002), 343.

[35] C. Hilbich, B. Kisters-Woike, J. Reed, C. L. Masters and K. Beyreuther, *J Mol Biol* 218 (1991), 149.
[36] M. O. Chaney, S. D. Webster, Y. M. Kuo and A. E. Roher, *Protein Eng* 11 (1998), 761.
[37] L. C. Serpell, C. C. Blake and P. E. Fraser, *Biochemistry* 39 (2000), 13269.
[38] J. T. Guo, R. Wetzel and Y. Xu, *Proteins* 57 (2004), 357.
[39] N. A. Whittemore, et al., *Biochemistry* 44 (2005), 4434.
[40] S. Shivaprasad and R. Wetzel, *Biochemistry* 43 (2004), 15310.
[41] R. Nelson, et al., *Nature* 435 (2005), 773.
[42] B. O'Nuallain and R. Wetzel, *Proc Natl Acad Sci U S A* 99 (2002), 1485.
[43] M. Dumoulin and C. M. Dobson, *Biochimie* 86 (2004), 589.
[44] A. C. Davis and M. J. Shulman, *Immunol Today* 10 (1989), 118.
[45] N. R. Whitelegg and A. R. Rees, *Protein Eng* 13 (2000), 819.
[46] Chemical Computing Group Inc., 1997.
[47] E. E. Abola, J. L. Sussman, J. Prilusky and N. O. Manning, *Methods Enzymol* 277 (1997), 556.
[48] J. L. Sussman, et al., *Acta Crystallogr D Biol Crystallogr* 54 (1998), 1078.
[49] A. C. Martin, *Proteins* 25 (1996), 130.
[50] S. F. Altschul, et al., *Nucleic Acids Res* 25 (1997), 3389.
[51] C. Chothia and A. M. Lesk, *J Mol Biol* 196 (1987), 901.
[52] T. T. Wu and E. A. Kabat, *Mol Immunol* 29 (1992), 1141.
[53] A. C. Martin, J. C. Cheetham and A. R. Rees, *Proc Natl Acad Sci U S A* 86 (1989), 9268.
[54] A. C. Martin, J. C. Cheetham and A. R. Rees, *Methods Enzymol* 203 (1991), 121.
[55] A. R. Rees, Searle,S.M.J., Henry,A.H.and Pedersen,J.T., *in* "Protein Structure Prediction" (S. M.J.E., ed.), p. 141–172. Oxford University Press, Oxford, 1996.
[56] J. T. Pedersen, Searle,S.M.J., Henry,A.H. and Rees,A.R., *Immunomethods* 1 (1992), 126.
[57] G. M. Morris, et al., *Journal of Computational Chemistry* 19 (1998), 1639.
[58] S. J. K. Weiner, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P., *J Am Chem Soc* 106 (1984), 765.
[59] S. J. K. Weiner, P. A.; Nguyen, D. T.; Case, D. A., *J Comput Chem* 7 (1986), 230
[60] Accelrys; MSI, San Diego, 2000.
[61] G. Murshudov, *Acta Crystallogr D Biol Crystallogr.* 53 (1997), 240.
[62] N. Collaborative Computational Project, *Acta Crystallogr D* 50 (1994), 760.
[63] C. Cole and J. Warwicker, *Protein Sci* 11 (2002), 2860.
[64] W. L. DeLano. DeLano Scientific, San Carlos, CA, USA, 2002.
[65] A. Szilagyi and P. Zavodszky, *Structure Fold Des* 8 (2000), 493.
[66] Y. Li, H. Li, S. J. Smith-Gill and R. A. Mariuzza, *Biochemistry* 39 (2000), 6296.
[67] C. G. Dealwis, et al., *J Mol Biol* 236 (1994), 342.
[68] V. M. Reyes, M. R. Sawaya, K. A. Brown and J. Kraut, *Biochemistry* 34 (1995), 2710.

# SNAPSHOTS OF THE PANTOTHENATE SYNTHETASE FROM MYCOBACTERIUM TUBERCULOSIS ALONG THE REACTION COORDINATE

Shuishu Wang
*Public Health Research Institute, 225 Warren Street, Newark, New Jersey 07103*

David Eisenberg
*Howard Hughes Medical Institute, UCLA-DOE Institute of Genomics and Proteomics, Molecular Biology Institute, UCLA, Box 951570, Los Angeles, CA 90095-1570*

## 1. ABSTRACT

Pantothenate synthetase (PS) from *Mycobacterium tuberculosis* represents a potential target for developing new anti-tuberculosis drugs. PS catalyzes the ATP-dependent condensation of pantoate and β-alanine to form pantothenate. We determined the crystal structures of PS from *M. tuberculosis* and its complexes with substrates ATP, pantoate, and β-alanine, as well as an ATP analog AMPCPP, a reaction intermediate pantoyl adenylate and a reaction product AMP, with resolutions from 1.6 to 2 Å. The PS structure reveals a dimer, and each subunit has two domains with tight association between domains. The active site cavity is on the N-terminal domain, covered by the C-terminal domain. There is a flexible loop that forms one wall of the active site cavity, and it opens and closes the active site cavity. The enzyme binds ATP and pantoate tightly in the active site, and brings the carboxyl oxygen of pantoate near the α-phosphorus atom of ATP for an in-line nucleophilic attack. The pantoyl adenylate thus formed is stabilized by tight binding interactions with enzyme active site residues, and protected from hydrolysis by closing of the active site cavity with the flexible loop becoming ordered. Binding of β-alanine can occur only after formation of the pantoyl adenylate intermediate. Crystal structures of these complexes provide a step-by-step view of the PS catalyzed reaction, and allow us to construct transient reaction intermediates and deduce detailed reaction mechanism. PS catalyzes the reaction by stabilizing the reaction intermediates along the reaction coordinate, and structural models of these intermediates can provide basis for inhibitor design.

## 2. INTRODUCTION

Pantothenate (vitamin B5) is an essential precursor for the biosynthesis of coenzyme A and acyl carrier proteins, both of which play critical roles in many cellular processes including energy metabolism and fatty acid metabolism [1]. Microorganisms and plants can synthesize pantothenate, while animals obtain this essential nutrient from their diet [2]. Therefore, the pantothenate biosynthetic pathway offers targets for developing drugs against microbial pathogens. Recently, Jacobs and coauthors [3] reported that a *Mycobacterium tuberculosis* (MTB) mutant defective in the de novo biosynthesis of pantothenate is highly attenuated in both immunocompromised and immunocompetent mice. This observation indicates that a functional pantothenate biosynthetic pathway is essential for virulence of MTB, and thus represents a new target for anti-tuberculosis agents.

The pantothenate biosynthetic pathway in bacteria comprises four steps catalyzed by enzymes encoded by the *panB*, *panC*, *panD* and *panE* genes [4]. The *panC* gene encodes a pantothenate synthetase (PS), which catalyzes the last step of pantothenate biosynthesis, the ATP-dependent condensation of pantoate and β-alanine to form pantothenate. The *panC* gene product in several organisms has been identified and characterized, including those in *Escherichia coli* [5], *Saccharomyces cerevisiae*, higher plants *Oryza sativa* and *Lotus japonicus* [6], fungus *Fusarium oxysporum* [7], and *M. tuberculosis* [8]. The PS enzymes from *E. coli*, higher plants, and MTB form dimers in solution.

The *E. coli* PS enzyme structure belongs to the cytidylyltransferase superfamily [5]. It has two distinct domains, a large N-terminal domain having a Rossmann fold and a smaller C-terminal domain containing a helical layer above a three-stranded antiparallel β-sheet. Based on structural comparison of the *E. coli* PS with other members of cytidylyltransferase superfamily having known structures, von Delft et al. [5] deduced the ATP and pantoate binding sites of the *E. coli* PS and proposed a hinged domain mechanism for opening and closing of the enzyme active site cavity.

Kinetic analysis of the MTB PS suggests that the enzyme-catalyzed reaction proceeds through two steps: the formation of an enzyme bound intermediate, pantoyl adenylate, from ATP and pantoate, followed by nucleophilic attack on the intermediate by β-alanine to form pantothenate and AMP [8]. The existence of pantoyl adenylate as an enzyme bound intermediate is suggested by the transfer of the $^{18}O$ label from the carboxyl group of pantoate to the reaction product AMP. However, pantoyl adenylate cannot be directly isolated or identified by TLC methods, presumably due to rapid lactonization [9].

Here, we present the crystal structure of the MTB PS enzyme and structures of its complexes with substrates, a substrate analog (AMPCPP), a reaction product (AMP), and a pantoyl adenylate reaction intermediate [10, 11]. The MTB PS has the same fold as the *E. coli* enzyme. However, in the MTB PS structure the domains of each subunit have a closed conformation, in contrast to those of the *E. coli* PS structure [5], and there is no significant movement between domains among all structures of the apo-enzyme and various complexes. A flexible region, which forms a wall of the active site cavity, becomes ordered in the reaction intermediate complex and closes the active site cavity, thus acting as a gate to the active site cavity of the MTB PS enzyme. The structure of the β-alanine complex gives direct evidence that the binding site for this last substrate exists only after formation of the pantoyl adenylate intermediate. Together these crystal structures lead to a detailed view of the overall reaction mechanism. Models of the transient reaction intermediate can be easily constructed from the experimentally observed complexes.

## 3. MATERIALS AND METHODS

### 3.1 Protein Production

The detailed procedures for cloning of the *panC* gene, expression and purification of the protein have been described [10]. Briefly, the MTB *panC* gene (Rv3602c) encoding the pantothenate synthetase was amplified from the genomic DNA of MTB strain H37Rv and inserted into a pET30a plasmid (Novagen), which produces a recombinant protein with an N-terminal 6xHis tag that can be cleaved off with enterokinase. BL21(DE3) cells containing the pET30-*panC* plasmid were grown in LB medium containing 50 μg/ml kanamycin at 37 °C and induced with 0.4 mM IPTG for 3 hours. Cells were collected and lysed by lysozyme treatment followed by sonication in 20 mM HEPES pH 7.8, 500 mM NaCl, and 0.5 mM PMSF. The protein was purified from a $Ni^{2+}$-charged HiTrap chelating column (Amersham) with 20 mM HEPES pH 7.8, 500 mM NaCl and a linear gradient of imidazole. The PS protein was subjected to enterokinase (New England Biolabs) digestion to cleave off the N-terminal fusion tag. Electrospray mass spectroscopy indicated that 9 residues from the C-terminus of the protein were also cleaved off by enterokinase digestion. However, enzyme activity assays suggested that this C-terminal truncation does not have any detectable effect on the enzyme activity [10]. The enterokinase digested protein was further purified with a second round of chromatography on a $Ni^{2+}$-HiTrap column followed by Superdex 75 column (Amersham) gel filtration.

### 3.2 Crystallization and Data Collection

Crystallization was carried out as described previously [10]. Crystals in space group $P2_1$ were obtained from drops set up with well solutions containing 10-15% PEG 3000, 5% glycerol, 2% ethanol, 20 mM $MgCl_2$, 150 mM $Li_2SO_4$ and 100 mM imidazole pH 8.0 at 20 ºC. Those in space group C2 were from similar conditions, except with 2% isopropanol in place of ethanol and with 200 mM $Li_2SO_4$. Cocrystallization with β-alanine was done with a well solution of 15% PEG 3000, 2% ethanol, 20 mM $MgCl_2$, 100 mM imidazole pH 8.0 and 20 mM β-alanine, from which crystals in space group $P2_12_12_1$ were obtained. However, these crystals do not have β-alanine in the active site. Cocrystallization with both ATP and pantoate was carried out in the same crystallization condition of the $P2_1$ crystals but with both ATP and pantoate added, from which isomorphous $P2_1$ crystals were obtained that have a pantoyl adenylate intermediate in the active site.

Crystals in complex with substrates were obtained by soaking apo-enzyme crystals in solutions containing substrates or the ATP analog AMPCPP. Soaking experiments were carried out by adding solutions containing substrates or AMPCPP directly to the drops, or to drops of pseudo mother liquor with transferred crystals, and incubating overnight or longer. Crystals were soaked in solutions containing ~10 mM each of AMPCPP alone, both AMPCPP and pantoate, both pantoate and β-alanine, both ATP and pantoate, and both AMP and β-alanine. Crystals were also soaked in solutions containing up to 20 mM pantothenate, but no pantothenate was found in the crystal structure.

Before data collection, crystals were soaked for 2-5 min in a cryogenic solution similar to well solutions or substrate soaking solutions with glycerol added to 30%, and they were then flash-frozen in a cryo stream of $N_2$ gas at 100 K. Diffraction data were collected at 100 K on a Rigaku FRD generator with an R-AXIS $IV^{++}$ detector. Data reduction and scaling were carried out with the programs DENZO and SCALEPACK [12]. Data processing statistics are reported in references [10] and [11]. Table 1 summarizes a few of the data sets, of which the structures are mentioned in this paper. The $P2_1$ crystal has two PS molecules per asymmetric unit; while the C2 crystal contains one molecule per asymmetric unit. In the C2 crystal, the active site is locked in an inactive conformation by the crystal packing (see below), and the pantoate and β-alanine binding sites are compromised. Therefore, only AMP was found when we soaked the C2 crystals with both AMP and β-alanine (data set 8, in Table 1), and those soaked with both ATP and pantoate gave an ATP complex (data set 9) instead of a reaction intermediate complex.

**TABLE 1**

Summary of crystal structure data and experiments

| | Complex | Experiment | SG | #/A.U. | resolution | PDB ID | Ref. |
|---|---|---|---|---|---|---|---|
| 1 | Apo enzyme | | $P2_1$ | 2 | 1.6 Å | 1MOP | [10] |
| 2 | AMPCPP and pantoate in A; AMPCPP in B | AMPCPP, pantoate, soak | $P2_1$ | 2 | 1.6 Å | 1N2E | [10] |
| 3 | Pantoyl adenylate | ATP, pantoate, cocrystal | $P2_1$ | 2 | 2.0 Å | 1N2H | [10] |
| 4 | Pantoyl adenylate | ATP, pantoate, soak | $P2_1$ | 2 | 1.7 Å | 1N2I | [10] |
| 5 | Pantoate | Pantoate, β-alanine, soak | $P2_1$ | 2 | 1.8 Å | 1N2J | [10] |
| 6 | AMP, β-alanine in A; AMP only in B | AMP, β-alanine soak | $P2_1$ | 2 | 1.85 Å | 1A86 | [11] |
| 7 | Apo enzyme | | C2 | 1 | 1.7 Å | 2A88 | [11] |
| 8 | AMP | AMP, β-alanine soak | C2 | 1 | 1.7 Å | 2A7X | [11] |
| 9 | ATP | ATP, pantoate, soak | C2 | 1 | 1.6 Å | 2A84 | [11] |

### 3.3 Structural Determination and Refinement

The crystal structure of the apo-enzyme in the $P2_1$ crystal was determined [10] by the molecular replacement method using AMORE [13], with subunit A of the *E. coli* PS structure (PDB ID code 1IHO [5]) as a model. The two domains were separated as two independent search models. The positions of the two N-terminal domains and one C-terminal domain were found from the rotation and translation searches, and the other C-terminal domain was generated by a non-crystallographic-symmetry (NCS) operation on the position of the known C-terminal domain with the NCS symmetry operators derived from the two N-terminal domains. The structure of the apo-enzyme was used to calculate phases for data sets of various complexes by rigid body refinement into new data sets for isomorphous crystals, or by molecular replacement with AMORE. The final refinement statistics are reported in references [10, 11]. Structural models were refined against diffraction data using CNS [14] and/or REFMAC [15]. After each cycle of refinement, models were manually adjusted with electron density maps using O [16]. The same subsets of data for $R_{free}$ calculation were kept between the CNS and REFMAC programs.

Besides the 9 residues that were digested off by enterokinase, there are 10 more residues from the C-terminus that do not have clear electron density. Based on its susceptibility to protease digestion and lack of clear electron density, it is reasonable to assume that these residues at the C-terminus of the protein are flexible, and do not have a defined structure in solution. It is clear from the structure that both N- and C-termini are away from the active site cavity, and therefore are unlikely to affect the enzyme catalytic activity. This is consistent with the results of enzyme activity assays [10].

## 4. STRUCTURE OF THE MTB PANTOTHENATE SYNTHETASE

### 4.1 Dimer Structure

The crystal structure of the MTB PS enzyme indicates that it is a dimer (**Figure 1**), resembling a butterfly when viewed from one side. The dimer interface is extensive, having a buried area of about 2150 $Å^2$. A cluster of 8 hydrophobic side chains at the center of the dimer interface, Phe174, Leu177, Leu144, and Val118, form the core of the dimer interaction. This cluster is surrounded by hydrogen bonds and salt bridges, as well as the specific β strand

interactions of the two-stranded intersubunit β sheet that is composed of the strand β5 from each subunit. The MTB PS enzyme also exists in solution as a dimer [8, 10], as do PS enzymes from higher plants [6, 7].

The two subunits of the dimer in the $P2_1$ crystal are similar to each other, with an rmsd of 0.9 for all Cα atoms except the disordered residues. Most of the deviation lies at residues 74 to 88, and residues 260 to 265, where the crystal packing environments are different. Residues 74 to 88 (β3-$3_{10}$3-α3'-α3) form one wall of the active site cavity (see **Figure 2** and the active site cavity section b 31 In subunit A these residues are ordered with residues 74 to 76 having high B factors, while in subunit B re- 75 to 80 (helix α3'and nearby residues) do not have electron density, and the rest of the residues up to residue 88 are partially disordered with high B factors. This flexible wall of the active site cavity becomes ordered in the complex with the pantoyl adenylate reaction intermediate, suggesting their function as a gate to the active site cavity (see below pantoyl adenylate binding interactions).

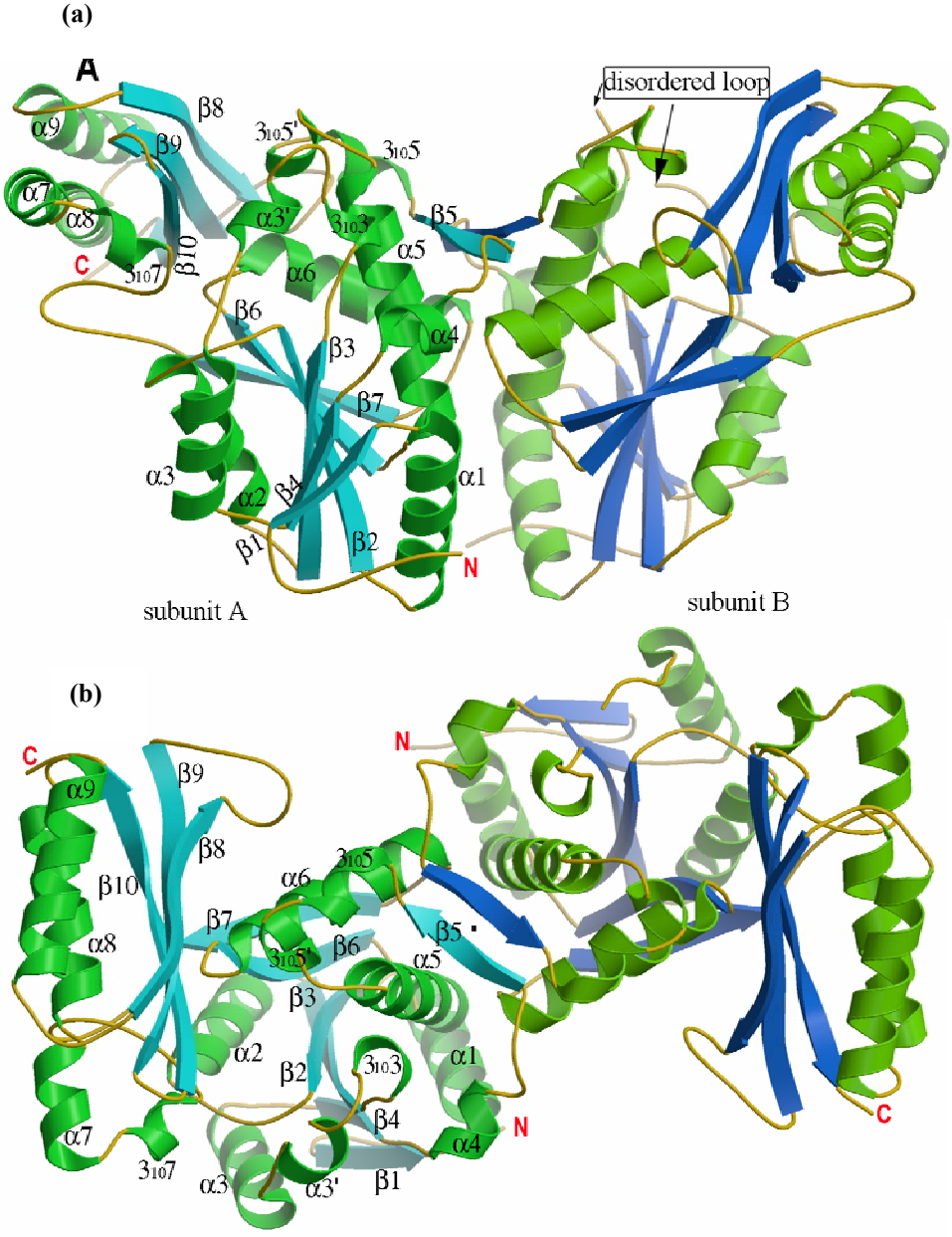



**Figure 1.** Ribbon diagram of the *M. tuberculosis* pantothenate synthetase dimer. **(a)** A side view of the dimer structure shows that it resembles the shape of a butterfly. **(b)** An orthogonal view of (A) from top, with the two-fold NCS symmetry axis (labeled with a dot) approximately perpendicular to the paper plane. Secondary structure elements for subunit A (left) are labeled. Those for subunit B are identical except that the short helix α3' is disordered. The figure was prepared from the apo-enzyme structure of the $P2_1$ crystal, with Molscript [17] and Raster3D [18].

The C2 crystal differs from the $P2_1$ crystal in that there is one polypeptide chain per asymmetric unit [11]. However, the protein molecule in the C2 crystal is also present as a dimer that is essentially identical to that in the $P2_1$ crystal, but the two-fold axis of the dimer coincides with one crystal symmetry axis. In the C2 crystal, a larger portion of the flexible wall, from residue 76 to 86, is completely disordered. The first turn of helix α3 unwinds, and helix $3_{10}3$, which contains residue Glu72 that is important for binding pantoate (see below pantoate binding site), also unwinds and forms an extended structure. This 32 ral difference is induced by crystal packing and the intrinsic flexibility of the loop.

### 4.2 Domain Interactions

Each subunit of the dimer has two well-defined domains (**Figure 1**). The N-terminal domain has a Rossmann fold, with a central parallel β sheet composed of strands β1 to β4, β6, and β7, and helices on both sides of the β-sheet. Strand β7 leads to the C-terminal domain, which starts with a hairpin loop followed by helix $3_{10}7$. The rest of the C-terminal domain has a simple two-layer structure: a layer of helices α7, α8, and α9 followed by a layer of antiparallel β sheet composed of strands β8 to β10.

The two domains have extensive interactions. There is a hydrophobic core composed of residues Leu123, Leu127, Tyr162, Val166 in the N-terminal domain and residues Tyr249, Leu257, and Leu269 in the C-terminal domain. Surrounding this hydrophobic core are three salt bridges, Glu159 to Arg267, Glu128 to Arg278, Glu126 to Arg253, and 6 additional hydrogen bonds linking the two domains. In addition, there are several water-mediated hydrogen bonding interactions. The relative positions of the two domains stay the same for the two subunits in the dimer and for all crystals including the $P2_1$, $P2_12_12_1$, and C2 crystals. These structural features suggest that domain hinge motion is unlikely to be the mechanism of opening and closing the active site cavity for MTB PS.

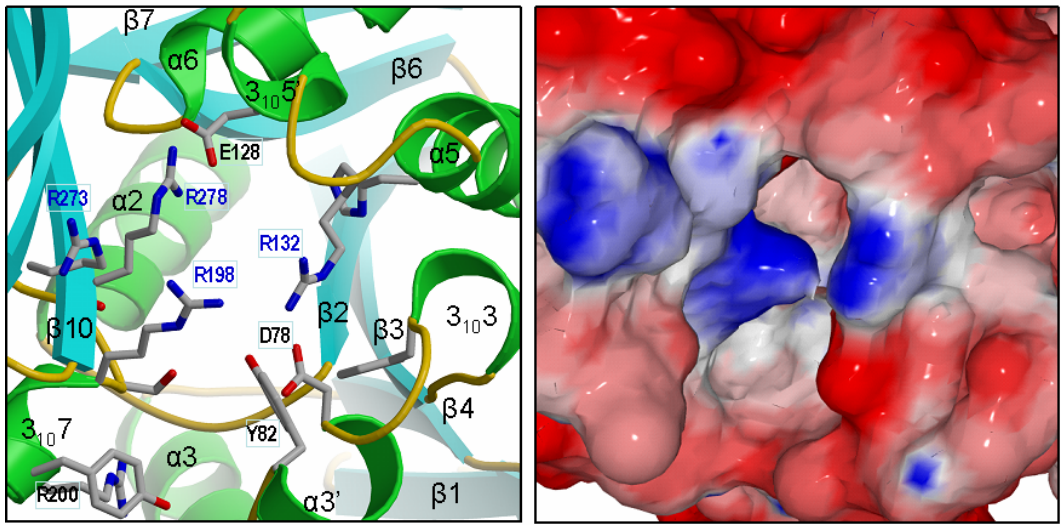



**Figure 2.** Active site cavity and the electrostatic potential surface around the entrance to the active site cavity. The figure was prepared from the subunit A of the $P2_1$ crystal. The two panels are approximately the same view. In the left panel, side chains at the entrance to the activity are shown as sticks. The segment from the end of β3 to the beginning of α3 forms the flexible wall, which is disordered in subunit B and in the C2 crystal thus opening the active site cavity.

### 4.3 Active Site Cavity

Typical for nucleotide binding (Rossmann fold) proteins, the active site is located at the C-terminus of the central parallel β sheet. The cavity is large and deep, and its bottom is at a cleft between strands β2 and β6 (**Figure 2**). The segment between β3 and α3 (loop-$3_{10}$3-loop-α3'-loop region) forms one side of the walls surrounding the cavity in subunit A of the $P2_1$ crystal. Part of this loop is disordered in subunit B, thus opens up the active site cavity. Other walls of the cavity are from β2-loop-α2, $3_{10}5'$-loop-α5, β6-loop-α6, and the loop after β7. Helix $3_{10}7$ and the β strands from the C-terminal domain partially cover the top of the active site cavity. The bottom of the active site cavity is mainly hydrophobic, while the top half of the cavity has several charged residues, including His44 and His47 at the N-terminus of α2 (the HIGH motif [5, 19]), Lys160 and Asp161 at the N-terminus of α6, and Arg198 of $3_{10}7$ (KMSKS motif [5]). At the entrance to the active site cavity, several charged side chains stick to the center,

leaving only a small opening to the cavity (**Figure 2**, right panel). Four arginine side chains, including Arg198, form a patch of positive charges at the entrance. These positive charges may steer the negatively charged substrates into the active site cavity. As we will see later, Arg198 plays an important role in binding the phosphate groups of ATP.

### 4.4 ATP Binding Interactions

The active site is well designed for binding ATP. **Figure 3** shows the initial difference electron density for the ATP molecule in the active site of the C2 crystal. The ATP molecule has an average temperature factor of 17.3 $Å^2$, virtually identical to the average temperature factor of the surrounding protein atoms, indicating that it is at full occupancy. A magnesium ion was found binding to the phosphate groups of ATP. It has a nearly perfect octahedral coordination. Its distances to the O2α, O2β, and O1γ atoms of the phosphate groups are 2.26 Å, 1.99 Å, and 2.12 Å, respectively; three other ligands are water molecules with distances of 1.96 Å, 2.11 Å, and 2.28 Å. The adenosine group of ATP binds tightly at the bottom of the active site cavity with hydrogen bonds and hydrophobic interactions [10, 11]. The phosphate groups stick toward the top of the active site cavity, and are located near the N-terminal ends of helices α2 and $3_{10}7$. Binding of the phosphate groups involves mainly main chain NH groups and charged side chains.

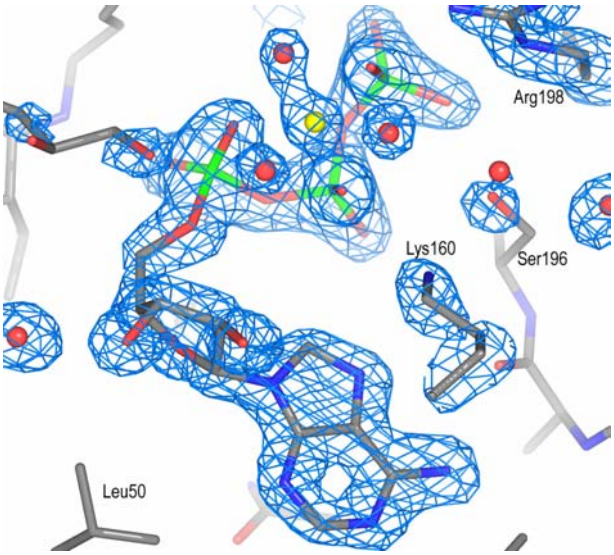



**Figure 3**. Initial Fo – Fc electron density in the active site of the C2 crystal soaked in a solution containing ATP and pantoate superimposed on the final refined model. The map was calculated at 1.6 Å and contoured at 3.5 σ. The ATP molecule has very well defined electron density. Side chains of Lys160, Ser196 and Arg198 moved relative to those in the apo-enzyme to interact with the phosphate groups, and thus have positive initial difference electron density. The figure is prepared with PYMOL [20].

Two His residues of the HIGH motif [5, 19, 21] interact with the phosphates of ATP. The first histidine, His44, has a hydrogen bond extending from its Nε2 to O1β of β-phosphate. The His47 side chain has a hydrogen bond from its Nε2 to the bridge oxygen between the α- and β-phosphate, suggesting that it functions as a general acid by donating a proton to the leaving pyrophosphate group. Both histidine residues have the Nδ1 atom forming a hydrogen bond to a main chain NH group, and thus the Nε2 atoms are hydrogen bond donors. Their imidazole rings are in close proximity to each other with the closest separation being ~3.2 Å, indicating favorable π-π interactions. These two side chains also play a role in binding the adenine group and stabilizing the reaction intermediate, because the imidazole rings are close to the adenine (closest distances of ~3.4 and ~4.3 Å for His44 and His47, respectively). Mutation of either histidine residue to alanine causes a greater than 1000-fold reduction in enzyme activity [22]. The Gly46 residue is highly conserved in this motif. It flanks one side of adenine (with Lys160 on the loop after β6 on the other side), where any side chain larger than a hydrogen atom would have steric hindrance. The KSMKS motif [23] is not conserved at the sequence level in PS enzymes, but it is evident from structural alignments [5]. Residues Ser196, Ser197, and Arg198, which are at the N-terminus of helix $3_{10}7$ and are part of the KSMKS motif, bind the β- and γ-phosphate groups. The side chains of Lys160 and Arg198 are partially disordered in the apo-enzyme structure, but become ordered and move to enable salt-bridges to the β- and γ-phosphate groups, respectively. The side chain of Lys160 also packs against one side of adenine. Mutation of this

residue to alanine reduces the enzyme activity more than 1000-fold, and it decreases affinity of the enzyme for ATP [22]. Another residue that binds ATP is Asp161, which has hydrogen bonds to two water ligands of the magnesium ion, and to O2* of the ribose group. Overall, the ATP molecule is rigidly held in the active site by many hydrophobic, π-electron, hydrogen bonding, and charge-charge interactions, which keep it in an optimal conformation for a nucleophilic attack by pantoate. Residues involved in binding the adenosine group are also important in stabilizing the pantoyl adenylate intermediate.

The binding interactions of ATP with protein are similar to those of AMPCPP [10]. However, the methylene group in AMPCPP replaces the bridge oxygen between the α- and β-phosphate groups of ATP, and thus it is not capable of forming a hydrogen bond with the Nε2 atom of His47. Consequently, the α-phosphate moves slightly away from amide nitrogen of Met40, and thus no hydrogen bond is formed between them in the AMPCPP complex. Therefore, ATP has slightly higher binding affinity to the enzyme active site than its non-reactive analog AMPCPP.

**4.5 Pantoate Binding Site**

In the $P2_1$ crystal that was soaked with pantoate and β-alanine, there is a full occupancy of pantoate in both active sites [10]. The binding site for pantoate is located in a pocket at the bottom of the active site cavity, next to the binding site of the adenosine group (**Figure 4a**). This is the same binding site proposed for the *E. coli* PS protein [5], although the pantoate molecule has a different conformation and therefore different binding interactions. The pantoate molecule is tightly bound at the bottom of the active site cavity, with side chains of two glutamine residues, Gln72 from helix $3_{10}3$ and Gln164 from helix α3, forming hydrogen bonds to its hydroxyl groups and one carboxyl oxygen. The two methyl groups face the hydrophobic groups at the bottom of the active site cavity, Pro38 and Phe157 side chains. The side chain of Met40 packs against the hydrophobic side of the pantoate molecule. Mutation of either glutamine residue to alanine greatly reduces the rate of pantoyl adenylate formation [22].

Soaking $P2_1$ crystals in solutions containing both AMPCPP and pantoate gave a full occupancy of AMPCPP, but with a glycerol in the pantoate binding site in subunit B, which has the flexible wall (residues 75-88) disordered. However, subunit A has a partial occupancy of both AMPCPP and pantoate, as indicated by their electron density and B factors. Increasing the concentration of pantoate in the soaking solution gave only a slightly higher occupancy of pantoate but a lower occupancy of AMPCPP [10]. All these observations indicate that AMPCPP and pantoate cannot coexist in the active site. This must be due to the steric and charge repulsions between the α-phosphate group and the carboxyl group of pantoate. Modeling both ATP and pantoate in one active site based on the fully occupied complexes will put one carboxyl oxygen atom of pantoate ~2.7 Å from the α-phosphorus atom of ATP [11]. This is essentially identical to what is observed in subunit A that has partial occupancy of both AMPCPP and pantoate. The carboxyl oxygen of pantoate is in a good position for an in-line nucleophilic attack on the α-phosphate. Because residues of the flexible wall (gate residues) are disordered in subunit B, the active site cavity is open and allows AMPCPP to diffuse easily into the active site. The extensive binding interactions of AMPCPP to the protein atoms make its binding thermodynamically more favorable than the binding of pantoate to the protein. Therefore, AMPCPP has a full occupancy in the active site cavity of subunit B of the complex with both AMPCPP and pantoate. On the other hand, the gate residues are less flexible in the subunit A due to crystal packing, thus making it kinetically less favorable for AMPCPP to bind, and allowing pantoate to have a partial occupancy in the active site of subunit A.

In the C2 crystal, the side chain of Gln72 on helix $3_{10}3$ swings out of the active site due to crystal packing, and the binding of pantoate is abolished. We observed only ATP in the active site with a full occupancy (see above ATP binding interactions) when we soaked the C2 crystal in a solution containing both ATP and pantoate [11]. There was no electron density for pantoate or pantoyl adenylate in the active site cavity. In the pantoate binding pocket, there was positive difference electron density, into which we modeled a glycerol molecule. This glycerol has a high temperature factor, suggesting that it has a low occupancy. Except for this Gln72 side chain, the rest of the binding site for pantoate is intact. In the $P2_1$ crystal, soaking with both ATP and pantoate results in a pantoyl adenylate intermediate in the active site cavity, and the disordered loop in subunit B becomes ordered [10] (see more details below). This indicates that loss of this glutamine side chain completely abolishes the enzyme catalytic activity for the first half of the reaction inside the crystal. The glutamine side chains are also involved in binding the reaction intermediate. Thus mutation of either residue could affect the stability of pantoyl adenylate intermediate. It is noteworthy that Asn69 was also found to be important for the enzyme activity [22]. This residue is not directly involved in binding substrates or the intermediate. However, it forms a hydrogen bond to the side chain of Gln72, and thus is important for keeping the Gln72 side chain in position. In addition, since both Asn69 and Gln72 are on

the one-turn $3_{10}3$ helix, mutation of Asn69 might also affect the stability of this $3_{10}$ helix and hence the position of the Gln72 side chain.

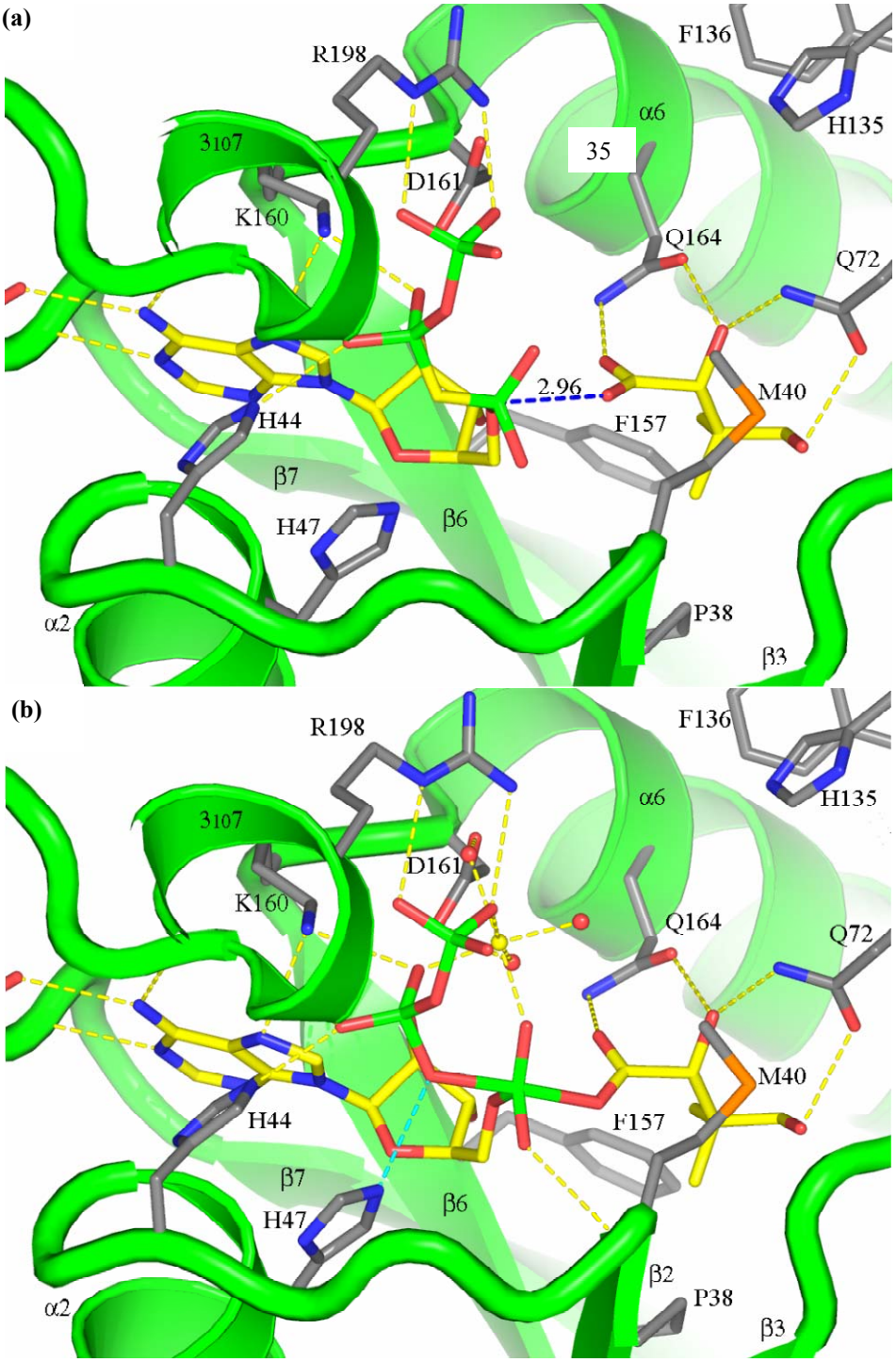



**Figure 4**. **(a)** Binding interactions of AMPCPP and pantoate in the active site of subunit A of the $P2_1$ crystal soaked in a solution containing both AMPCPP and pantoate. Both molecules have a partial occupancy (see text for details), and the magnesium ion is not well defined in this active site. However, binding position and interactions are essentially identical to those of AMPCPP and pantoate with a full occupancy when each is bound in the active site alone. The distance between the α-phosphorus atom and the carboxyl oxygen is ~2.96 Å, shown in blue dashed line. **(b)**

A model of the trigonal bipyramidal intermediate based on crystal structures of the ATP complex, AMPCPP complex and pantoate complex. Formation of the intermediate requires only the movement of the α-phosphorus atom towards the carboxyl oxygen. All favorable interactions with active residues are preserved. His47 is poised to function as a general acid.

### 4.6 Binding Interactions of the Reaction Intermediate, Pantoyl Adenylate

Soaking the $P2_1$ crystals with solutions containing both ATP and pantoate, or crystallization in the presence of both substrates, resulted in a reaction intermediate, pantoyl adenylate, in the active site [10]. Cocrystallization resulted in both active sites in the dimer being fully occupied with the pantoyl adenylate molecule. However, crystals soaked with ATP and pantoate have one active site (subunit B) fully occupied but the other partially occupied, indicated by their difference in B factors. Subunit A of the dimer also has a lower occupancy of AMPCPP when crystals were soaked in solutions containing AMPCPP. This is due to crystal packing, which makes the flexible wall of the active site cavity ($3_{10}$3-loop-α3'-loop region) ordered in subunit A and thus the active site less accessible for bulky molecules ATP and AMPCPP.

Pantoyl adenylate has extensive binding interactions with the active site residues (**Figure 5**). The molecule is almost linear and fits snugly in the bottom of the active site cavity. Both pantoate and adenylate moieties sit in the same positions in the active site as the pantoate molecule of the pantoate complex and the adenosine group of ATP in the ATP complex, and preserve all strong interactions with protein atoms. The tight binding of pantoyl adenylate stabilizes this highly reactive intermediate. One important feature of the pantoyl adenylate complex is that the disordered loop in subunit B becomes ordered, in both cocrystallization and soaking the apo-enzyme crystals with both ATP and pantoate. This suggests that the flexible wall can function as a gate, opening for substrates, especially the bulky ATP molecule, to bind and closing the active site cavity to protect the highly reactive pantoyl adenylate.

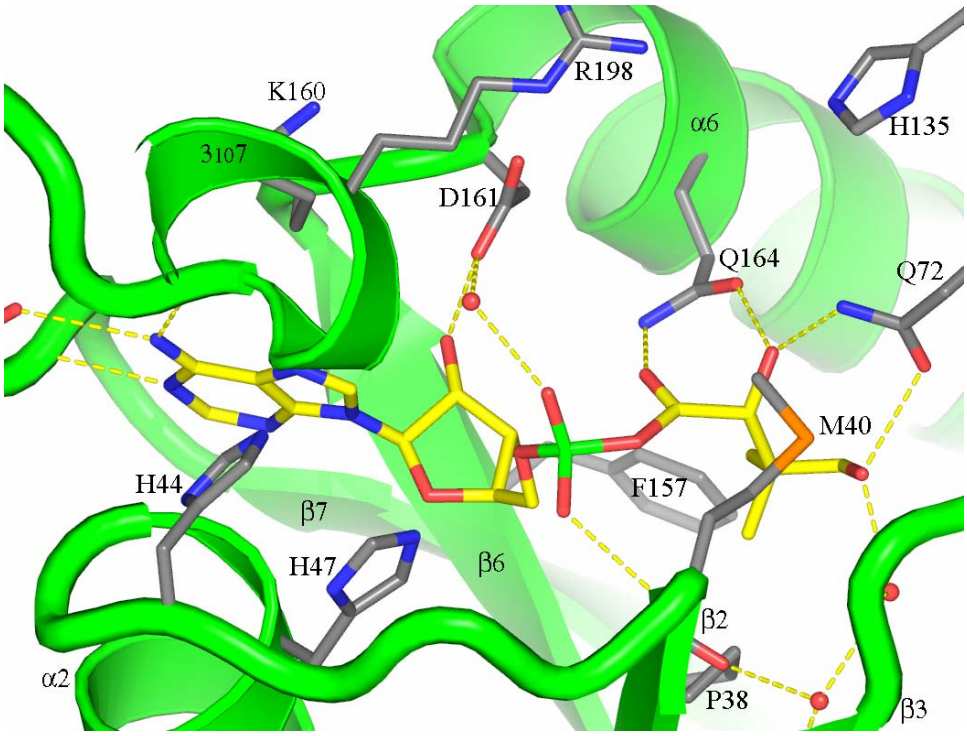



**Figure 5**. Binding interactions of the pantoyl adenylate intermediate. The intermediate binds tightly at the bottom of the active site cavity with many hydrogen bonds and hydrophobic interactions. Hydrogen bonds to the O3* of ribose and some water mediate hydrogen bonds are not shown for clarity.

### 4.7 Binding Interactions of Beta-Alanine

Cocrystallization and soaking crystals with β-alanine [10] failed to yield a complex with this substrate, suggesting that the binding site for β-alanine exists only after pantoyl adenylate is formed in the active site. Based on the structure of the pantoyl adenylate complex, it is likely that the phosphate group of pantoyl adenylate serves as an anchor for the initial binding of β-alanine by offering hydrogen-bonding partners and/or favorable charge-charge interactions. Furthermore, an AMP molecule bound at the active site may also assist the binding of β-alanine from

its phosphate group. Therefore, we soaked crystals of PS in a solution containing both β-alanine and AMP, and we obtained a crystal structure of the PS enzyme in complex with β-alanine and AMP from the $P2_1$ crystals [11].

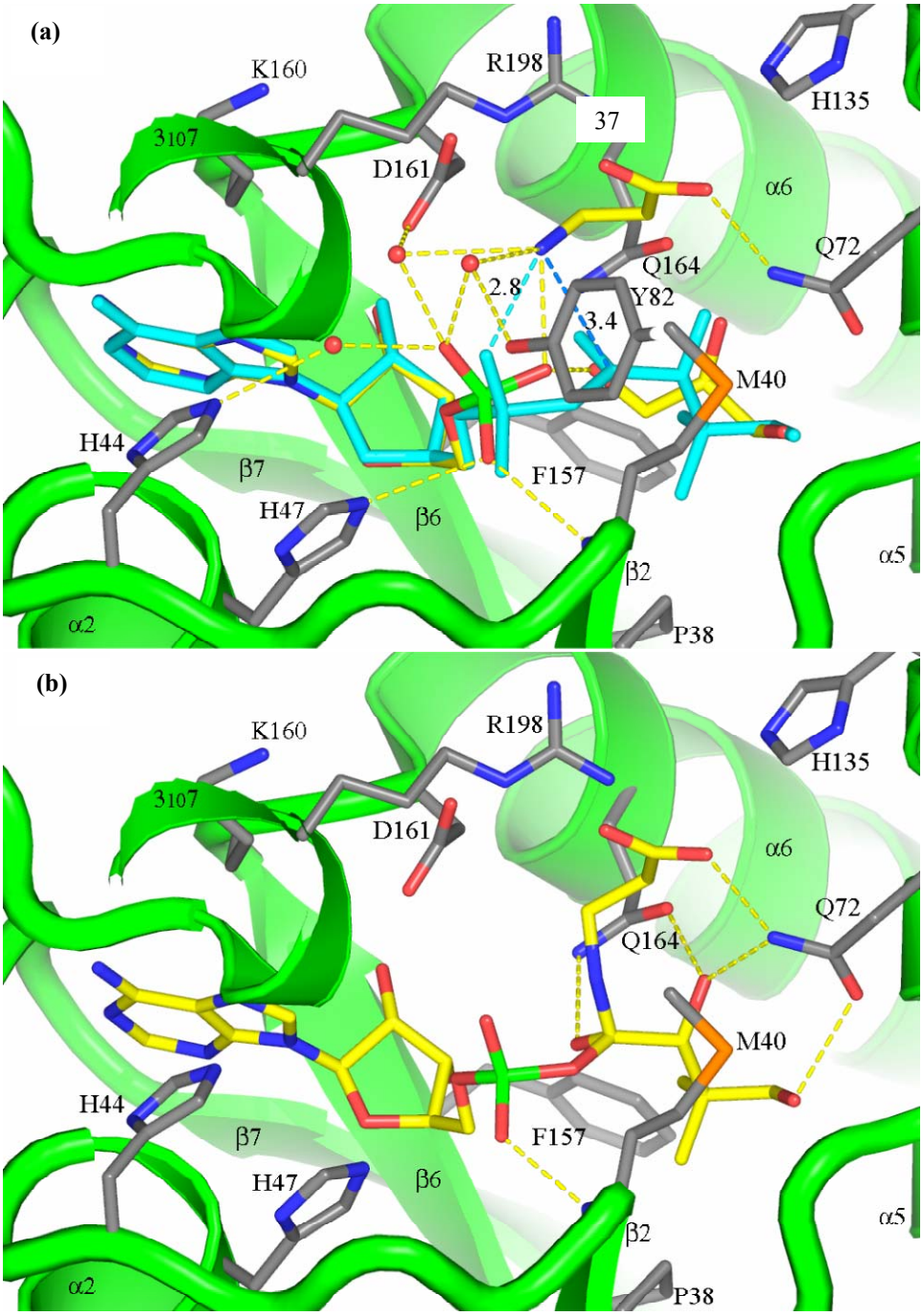



**Figure 6**. **(a)** Structural superposition of the active site of the pantoyl adenylate complex with that of the β-alanine/AMP complex. The active site residues align well with an rmsd of Cα less than 0.15 Å. The pantoyl adenylate molecule is colored in cyan. The protein structure shown in the figure is from the β-alanine/AMP complex. Hydrogen bonds to the phosphate group of AMP and to β-alanine are shown in yellow dashed lines. Some water mediated hydrogen bonds are not shown for clarity. The Tyr82 side chain is from a loop in the front that is clipped away to reveal the bound AMP and β-alanine. Based on the structural alignment, the amino group of β-alanine is ~2.8 Å from the phosphate oxygen of pantoyl adenylate, and ~3.4 Å from the carbonyl carbon, shown in blue dashed lines. **(b)** Model of the tetrahedral intermediate structure based on the structural superposition in (A). The Tyr82 side chain is not shown for clarity. A tetrahedral geometry can be obtained by a torsional rotation of the amino nitrogen of β-alanine to move it to ~1.5 Å above the carbonyl carbon. All favorable interactions with protein atoms are preserved, and thus the intermediate is stabilized by the enzyme.

The β-alanine molecule binds in the active site of subunit A in the upper part of the active site cavity, with its amino group near the phosphate group of AMP (**Figure 6a**). The amino group forms hydrogen bonds to the phosphate group of AMP and to two water molecules. Both water molecules are fixed in position through one hydrogen bond to the phosphate group and a second hydrogen bond to the side chain of Asp161 or Tyr82. The carboxyl group of β-alanine has one hydrogen bond to the Nε2 atom of th[illegible]72 side chain. This carboxyl is also close to the side chains of Arg198 and His135, with distances of ~4.0 [illegible] Å, respectively, indicating favorable charge-charge interactions and π-electron interactions. One side of β-alanine faces side chains of Met40 and Tyr82; the other side faces the large cavity of the active site, which is filled with a few ordered water molecules in the crystal structure. Binding of β-alanine is not as tight as that of other substrates and its binding site in the crystal is not fully occupied. However, the binding interactions limit molecules that can bind well and have a nucleophilic attack on the reaction intermediate. Therefore, structural analogs of β-alanine are all poor substrates [8].

In subunit B of the dimer, where residues 74 to 83 (including Tyr82) are disordered, only AMP but no β-alanine was found in the active site [11]. Also in this case, disorder of this active site cavity wall weakens binding interactions for β-alanine because its binding site is exposed to bulk solvent and thus the charge-charge interactions are dampened. Similarly, in the C2 crystal, which has one molecule per asymmetry unit and the flexible loop disordered, only AMP was found in the active site when crystals were soaked in a solution containing both AMP and β-alanine.

Superposition of the structure of the β-alanine/AMP complex with that of the pantoyl adenylate complex (**Figure 6a**) shows that the amino group of β-alanine was positioned above the carbonyl group of pantoyl adenylate such that the nitrogen atom was ~3.4 Å from the carbonyl carbon. The O1α atom of pantoyl adenylate can form a hydrogen bond with the amino group of β-alanine, serving as an anchor for the initial binding of β-alanine. With changes in only one torsion angle, the amino group of β-alanine can move towards the carbonyl carbon for a nucleophilic attack and attain a nearly perfect geometry for a tetrahedral intermediate (**Figure 6b**).

### 4.8 AMP Binding Interactions

As described above, we obtained AMP complexes from both $P2_1$ and C2 crystals by soaking crystals in a solution containing both AMP and β-alanine. Binding interactions of AMP with the active site residues in these complexes are essentially identical. The adenosine group of AMP binds in the same position as that of the pantoyl adenylate, AMPCPP, and ATP: it fits snugly in its binding pocket at the bottom of the active site cavity through hydrophobic and hydrogen bonding interactions. The adenine group is flanked by Gly46 on helix α2 and Lys160 on the loop after β6. Its N1 and N6 atoms have hydrogen bonds to main chain atoms. The hydroxyl groups of ribose form hydrogen bonds with the Asp161 side chain and a few main chain atoms at the bottom of the active site cavity. The phosphate group, however, has torsional flexibility, and it rotates slightly relative to the α-phosphate of ATP, allowing one of its oxygen atoms to form hydrogen bonds to the side chain of His47 and to the amide nitrogen of Met40 simultaneously (**Figure 6a**).

## 5. OVERALL ENZYME-CATALYZED REACTION MECHANISM

Putting all structural data together allows us to have a detailed view of the overall enzyme-catalyzed reaction mechanism. When the active site is empty, the disordered loop opens the active site and thus allows the ATP molecule to come in and bind. Because ATP is relatively bulky and is held rigidly in the active site by many binding interactions, it is likely that ATP binds first. The smaller sized pantoate then comes in, and initiates a nucleophilic attack on the α-phosphate while it binds in its binding pocket [8, 24]. When both ATP and pantoate are present in the same active site, the nucleophilic reaction must occur (**Figure 4a**). Formation of the transient trigonal bypyramidal intermediate involves only the movement of the phosphate atom towards the carboxyl oxygen of pantoate to be coplanar with the three equatorial oxygen atoms (**Figure 4b**). Although transient in nature, this intermediate is stabilized by the enzyme because all the favorable interactions are preserved, and the unfavorable steric hindrance and charge repulsions between ATP and pantoate are eliminated. The trigonal bipyramidal intermediate then dissociates, and simultaneously the pyrophosphate group leaves the active site. The magnesium ion and the positively charged side chains around the β- and γ-phosphate groups draw the negative charges towards the leaving pyrophosphate. His47 is also likely to facilitate the reaction by donating a proton to the leaving pyrophosphate. This leaves a reaction intermediate, pantoyl adenylate in the active site and completes the first half of the enzyme catalyzed reaction. Pantoyl adenylate is stabilized by many strong binding interactions, as we have seen earlier. It is also protected by the closing of the active site when the flexible loop becomes ordered. Pantoyl adenylate is stable only when tightly bound in the enzyme active site. In solution, it decomposes rapidly to give

pantoyl lactone and AMP [9]. The dissociation of the trigonal bipyramidal intermediate necessitates the simultaneous leaving of the pyrophosphate, because there would be unfavorable charge and steric repulsions once pyrophosphate is formed. This reaction, however, is reversible, as was demonstrated by positional isotope exchange experiments that $^{18}O$ labels scrambled when incubating [$\beta\gamma$-$^{18}O_6$]-ATP and pantoate with the MTB PS enzyme [24].

The presence of pantoyl adenylate intermediate and the ordering of the flexible wall create the binding site for β-alanine, the last substrate. The β-alanine molecule is relatively small, capable of passing through the small opening at the top of the active site cavity. The phosphate group of pantoyl adenylate can have a hydrogen bond to the amino group of β-alanine, thus serving as an initial anchor for binding β-alanine. The carboxylate group of β-alanine has favorable interactions with the side chains of Arg198 and His135 and a hydrogen bond to Gln72. There is a shift of the Arg198 side chain of ~1.8 Å towards the carboxylate group of β-alanine relative to that in the pantoyl adenylate complex. This binding position for β-alanine puts its amino group at a good position for a nucleophilic attack on the carbonyl carbon of the pantoyl group. The distance of the amino nitrogen is only ~3.4 Å from the carboxyl carbon of pantoyl adenylate. A model of the tetrahedral intermediate can be readily prepared from the crystal structures of the pantoyl adenylate complex and the β-alanine/AMP complex (**Figure 6b**). Formation of the tetrahedral intermediate requires only a torsional rotation of the amino nitrogen atom of β-alanine toward the carbonyl carbon of pantoyl adenylate. Similar to that of the trigonal bypyramidal intermediate, formation of the tetrahedral intermediate preserves all favorable binding interactions of active site residues with pantoyl adenylate and β-alanine. Therefore, the intermediate is stabilized by the enzyme. Dissociation of the tetrahedral intermediate then forms pantothenate and AMP. The planarity of the peptide bond of pantothenate causes significant rearrangement of the molecule, and several hydrogen bonds to active site residues are broken. In addition, the newly formed pantothenate has steric clashes and charge repulsions with AMP and some active site residues. Therefore, pantothenate must leave the active site once it is formed. The AMP molecule then diffuses away to regenerate the enzyme active site for next cycle of catalysis. Pantothenate has very low binding affinity to the enzyme. We have tried soaking and growing crystals in the presence of pantothenate, but we failed to obtain a complex with this compound. This is consistent with the finding that pantothenate is a poor inhibitor of the enzyme [8]. The AMP molecule, on the other hand, has good binding interactions in the active site. However, it typically has a low concentration inside cells, and thus can easily diffuse out of the active site. Moreover, ATP has many favorable interactions in the active site of the enzyme, and it can readily displace AMP.

## 6. REFERENCES


[1] S. Jackowski. Escherichia coli and Salmonella typhimurium: Cellular and molecular biology, American Society for Microbiology, Washington, DC 1996, pp. 687-694.

[2] W.K. Maas, Proc Int Congr Biochem 11 (1960) 161-168.

[3] V.K. Sambandamurthy, X. Wang, B. Chen, R.G. Russell, S. Derrick, F.M. Collins, S.L. Morris and W.R. Jacobs, Jr., Nat Med 8 (2002) 1171-4.

[4] W.K. Merkel and B.P. Nichols, FEMS Microbiol Lett 143 (1996) 247-52.

[5] F. von Delft, A. Lewendon, V. Dhanaraj, T.L. Blundell, C. Abell and A.G. Smith, Structure (Camb) 9 (2001) 439-50.

[6] U. Genschel, C.A. Powell, C. Abell and A.G. Smith, Biochem J 341 ( Pt 3) (1999) 669-78.

[7] A. Perez-Espinosa, T. Roldan-Arjona and M. Ruiz-Rubio, Mol Genet Genomics 265 (2001) 922-9.

[8] R. Zheng and J.S. Blanchard, Biochemistry 40 (2001) 12904-12.

[9] T. Wieland, W. Loewe, A. Kreiling and G. Pfleiderer, Biochem Z 339 (1963) 1-7.

[10] S. Wang and D. Eisenberg, Protein Sci 12 (2003) 1097-108.

[11] S. Wang and D. Eisenberg, Biochemistry (2005) In press.

[12] Z. Otwinowski and W. Minor, Methods Enzymol 276 (1996) 307-326.

[13] J. Navaza, Acta Crystallogr A 50 (1994) 157-163.

[14] A.T. Brunger, P.D. Adams, G.M. Clore, W.L. DeLano, P. Gros, R.W. Grosse-Kunstleve, J.S. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson and G.L. Warren, Acta Crystallogr D Biol Crystallogr 54 ( Pt 5) (1998) 905-21.

[15] G.N. Murshudov, A.A. Vagin, A. Lebedev, K.S. Wilson and E.J. Dodson, Acta Crystallogr D Biol Crystallogr 55 ( Pt 1) (1999) 247-55.

[16] T.A. Jones, J.Y. Zou, S.W. Cowan and Kjeldgaard, Acta Crystallogr A 47 ( Pt 2) (1991) 110-9.

[17] P.J. Kraulis, Journal of Applied Crystallography 24 (1991) 946-950.

[18] E.A. Merritt and M.E. Murphy, Acta Crystallogr D Biol Crystallogr 50 (1994) 869-73.

[19] D.G. Barker and G. Winter, FEBS Lett 145 (1982) 191-3.
[20] W.L. DeLano, The PyMOL molecular graphics system, DeLano Scientific, San Carlos, CA, 2002.
[21] P. Bork, L. Holm, E.V. Koonin and C. Sander, Proteins 22 (1995) 259-66.
[22] R. Zheng, T.K. Dam, C.F. Brewer and J.S. Blanchard, Biochemistry 43 (2004) 7171-8.
[23] C. Hountondji, P. Dessen and S. Blanquet, Bio 40 e 68 (1986) 1071-8.
[24] L. Williams, R. Zheng, J.S. Blanchard and F.M. Raushel, Biochemistry 42 (2003) 5108-13.