

# Structure-based Non-canonical Amino Acid Design to Covalently Crosslink an Antibody-Antigen Complex

Jianqing Xu,<sup>a§</sup> Drew Tack,<sup>b§</sup> Randall A. Hughes,<sup>b,c§</sup>  
Andrew D. Ellington,<sup>b,c\*</sup> and Jeffrey J. Gray<sup>a\*</sup>

a. Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD, 21218, USA

b. Department of Chemistry and Biochemistry, and Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin, Texas 78712, USA

c. Applied Research Laboratories, The University of Texas at Austin, Austin, TX 78758, USA

§ Authors contributed equally

---

\* Corresponding authors, E-mail: *andy.ellington@mail.utexas.edu, jgray@jhu.edu*

## **ABSTRACT**

Engineering antibodies to utilize non-canonical amino acids (NCAA) should greatly expand the utility of an already important biological reagent. In particular, introducing crosslinking reagents into antibody complementarity determining regions (CDRs) should provide a means to covalently crosslink residues at the antibody-antigen interface. Unfortunately, finding the optimum position for crosslinking two proteins is often a matter of iterative guessing, even when the interface is known in atomic detail. Computer-aided antibody design can potentially greatly restrict the number of variants that must be explored in order to identify successful crosslinking sites. We have therefore used Rosetta to guide the introduction of an oxidizable crosslinking NCAA, L-3,4-dihydroxyphenylalanine (L-DOPA), into the CDRs of the anti-protective antigen scFv antibody M18, and have measured crosslinking to its cognate antigen, domain 4 of the anthrax protective antigen. Computed crosslinking distance, solvent accessibility, and interface energetics were three factors considered that could impact the efficiency of L-DOPA-mediated crosslinking. In the end, **ten** variants were synthesized, and crosslinking efficiencies were generally 10% or higher, with the best variant crosslinking to 55% of the available antigen. The results suggest that computational analysis can be used in a pipeline for engineering crosslinking antibodies. The rules learned from L-DOPA crosslinking of antibodies may also be generalizable to the formation of other crosslinked interfaces and complexes.

### **Keywords:**

Computer-aided Design, Structure-based Design, Rosetta, Antibody, Non-canonical Amino Acid, Crosslinking, Binding Affinity, L-DOPA

## INTRODUCTION

Antibodies are key components of the immune system with broad diversity to recognize a variety of antigens. Antibody-based therapeutic, diagnostic, and industrial applications frequently require antibodies having high stability and strong binding affinity. With the development of computational techniques and a number of successful experiences in protein modeling and design (Lippow and Tidor, 2007; Mandell and Kortemme, 2009), computational antibody design has begun to play an important role in predicting improvements to antibody function. Computational design of antibodies has been used to enhance binding affinity (Barderas et al., 2008; Clark et al., 2006; Lippow et al., 2007), to improve stability by improvement of thermal/aggregation resistance (Chennamsetty et al., 2009; Miklos et al., 2012), and to alter binding specificity (Farady et al., 2009), and others (A. Caravella et al., 2010; Kuroda et al., 2012; Midelfort et al., 2004; Pantazes and Maranas, 2010).

To date, though, most computational design methods have focused on manipulating the twenty natural proteogenic amino acids to modify molecular forces such as electrostatics (Lippow et al., 2007), hydrophobic interactions (Chennamsetty et al., 2009), hydrogen bonds (Clark et al., 2006), and salt bridges (Miklos et al., 2012). However, recent advances in engineering the translation system have now allowed for the site-specific insertion of non-canonical amino acids (NCAAs) with a variety of functionalities into proteins with good efficiency (Wang et al., 2006; Xie and Schultz, 2006). Such NCAAs can be used to improve the stability and pharmacokinetics of therapeutic proteins (Cho et al., 2011), to augment binding (Liu et al., 2009), and to provide a myriad of chemical handles to study protein structure and function (Jones et al., 2010; Tsao et al., 2006; Zhang et al., 2002).

The generation of protein-protein crosslinks by inserting NCAAs into proteins could prove useful for a variety of applications. [To this end, a number of crosslinking-capable NCAAs have been incorporated into proteins in a site-specific manner utilizing an array of functionalized amino acids. These crosslinking functionalities include photo-crosslinkable aryl-azides \(Chin et al., 2002b\), benzophenones \(Chin et al., 2002a\) and diazirines \(Ai et al., 2011\) as well as the oxidizable crosslinker, L-DOPA \(Alfonta et al., 2003\). While any of the crosslinkers might benefit from a quantitative placement methodology, we chose L-DOPA because the periodate](#)

induced oxidation allowed for more control over the crosslinking conditions relative to photo-inducible crosslinkers that have been found to spuriously crosslink during sample handling (Chin et al., 2002b). In addition, the nucleophile-driven cross-linking mechanism of L-DOPA has been extensively characterized with a variety of proteinaceous nucleophiles (Liu et al., 2006).

L-DOPA has previously been used to successfully crosslink the monomeric domains of a dimeric sortase A for structural studies (Umeda et al., 2009), to enhance the affinity of low-affinity peptide probes for a kinase SH3 bioassay (Umeda et al., 2010), and to site-specifically label proteins with polysaccharides (Ayyadurai et al., 2011). While previously reported uses of L-DOPA as a site-specific crosslinker have yielded examples of effective crosslinking (as shown by SDS-PAGE or Western blot analyses), the actual efficiencies of crosslinking have never been reported (Burdine et al., 2004; Umeda et al., 2009; Umeda et al., 2010). These previous reports indicated that it was possible to place L-DOPA by intuition, but did not provide more quantitative assessments of what parameters impacted crosslinking efficiency (Umeda et al., 2009; Umeda et al., 2010).

In this paper, we explore how the Rosetta suite of computational protein design tools might be used to predict the site-specific, functional incorporation of L-DOPA into an antibody, allowing it to crosslink to its cognate antigen. A better understanding of where and how to insert crosslinking moieties into an antibody combining site could lead to the development of tools for validating antibody-antigen structural models (Pimenova et al., 2008) and to reagents capable of binding analytes with extremely high affinities and specificities.

As a proof-of principle demonstration, we chose a complex with a known structure, the anti-anthrax antibody M18 bound to anthrax protective antigen (PA) (Leysath et al., 2009). PA is a component of the tripartite toxin secreted by *Bacillus anthracis*, binds to cellular receptors, and assists host cellular targeting and transport of the lethal factor (LF) and edema factor (EF) into cytoplasm (Moayeri and Leppla, 2004; Young and Collier, 2007). M18 is a neutralizing antibody (Leysath et al., 2009) derived by directed evolution from monoclonal antibody 14B7 (Harvey et al., 2004; Little et al., 1988), binds to the fourth domain of PA (PAD4), and effectively blocks PA binding to cellular receptors such as CMG2 to mitigate anthrax toxicity.

## METHODOLOGY

### Computational Methods

#### *Creation of models of L-DOPA antibody mutants in complex with antigen*

Models for various mutants of the antibody-antigen complex were created using Rosetta (Leaver-Fay et al., 2011) with L-DOPA placed in various positions within the antibody paratope. Coordinates for the wild-type M18-PAD4 complex were downloaded from the Protein Data Bank (Berman et al., 2000) (PDB ID 3ETB). To remove crystal packing effects and obtain a Rosetta-minimized reference structure, fixed-backbone side-chain packing and minimization (1000 decoys) on the entire protein complex was performed using Rosetta's score12. The lowest-scoring structure was used for the calculations for the predictive introduction of L-DOPA into the CDRs of M18. The Rosetta 3.4 (revision 51671, available at [www.rosettacommons.org](http://www.rosettacommons.org)) command line used to run the "fix\_bb" protocol was:

```
fixbb.linuxgccrelease -s Crystal.pdb -nstruct 1000 -use_input_sc
    -minimize_sidechains -run:multiple_processes_writing_to_one_directory
    -packing:repack_only -ex1 -ex2aro -database $rosetta/rosetta_database
```

For each interface Lys on the antigen, neighboring antibody residues within 10 Å ( $C_{\beta}$ - $C_{\beta}$  distance) were selected as potential mutation sites. Each antibody residue within the cutoff distance was substituted to L-DOPA individually, followed by fixed-backbone side-chain packing (20 decoys) of the nearby residues (<10 Å) to accommodate the local changes. For these and any further calculations where L-DOPA is present, Rosetta uses the molecular mechanics-based scoring function (mm\_std) and associated NCAA rotamer library (Renfrew et al., 2012).

To carry out these calculations, the position of the L-DOPA mutation and the positions of the neighboring residues were specified in a "resfile", and the same "fixbb" protocol read the "resfile" and substituted the target residue to L-DOPA, followed by side chain repacking including all the neighboring residues. A Rosetta 3.4 (revision 51671) command line example is:

```
fixbb.linuxgccrelease -s Best_Prepacked.pdb -nstruct 20 -use_input_sc
    -resfile 315_LYS_J_679--139_SER_H_31.resfile -score:weights mm_std
    -minimize_sidechains -ex1 -ex2 -database $rosetta/rosetta_database
```

### *Model relaxation with crosslink constraint*

Some measures are performed on a structure where the L-DOPA is artificially constrained to be proximal to the target lysine residue. For these calculations, we employed an [empirically-determined](#) linear constraint potential,

$$E_{\text{constr}} = -100 + 400 \times |d_{\text{XL}} - 3.5|$$

where  $d_{\text{XL}}$  is the distance in Ångstroms between  $C_{\gamma}$  atom on the L-DOPA ring (the atom bound to the  $C_{\beta}$  atom) and the Lys  $N_{\epsilon}$  atom, and  $E_{\text{constr}}$  is the constraint energy in Rosetta Energy Units (REU). [The constraint weights were chosen to bring the L-DOPA and Lys into proximity, in order to evaluate interface compatibility. This constraint energy was not included in the final calculated interface score.](#) All the neighboring residues within 10 Å ( $C_{\beta}$ - $C_{\beta}$  distance) of the L-DOPA/Lys pair were repacked to accommodate the constraint. The constrained conformation was generated using a custom PyRosetta script with PyRosetta 2.012, revision 51671 (PyRosetta available at [www.rosettacommons.org](http://www.rosettacommons.org), script available upon request).

### *Crosslinking distance*

After selecting the L-DOPA position and the target Lys, all the distances of potential crosslinking atom pairs (lysine  $N_{\epsilon}$  atom and L-DOPA atoms  $C_2$ ,  $C_5$ , and  $C_6$ ) were evaluated, and the one with the minimum value represented the crosslinking distance ( $d_{\text{XL}}$ ).

### *Solvent-accessibility measures*

As a proxy for solvent accessibility, the number of atoms within 3 Å of the L-DOPA side-chain atoms was computed (Simons et al., 1999). All non-L-DOPA atoms were considered, both from the antibody and antigen, and including hydrogens (as placed by Rosetta). A second solvent accessibility measure is the solvent accessible surface area (SASA) (Le Grand and Merz, 1993), calculated using an embedded function in PyRosetta 2.012 using a probe radius of 1.4 Å.

### *Rosetta energy calculations*

The interface score  $I_{\text{sc}}$  is defined as

$$I_{\text{sc}} = S_{\text{bound}} - S_{\text{unbound}}$$

where  $S_{\text{bound}}$  and  $S_{\text{unbound}}$  represent the total scores of the antibody-antigen complex in bound and unbound form, respectively. The unbound form was scored after separating the antibody and antigen to a very large distance without additional side chain repacking. Scores are given in Rosetta Energy Units (REU), which approximate kcal/mol.

After repacking side chains under a constraint on the L-DOPA/Lys pair and, the new pose of the M18/PAD4 complex was saved in a separate file. The constrained interface score,  $I_{sc}^{\text{constr}}$ , was then obtained by rescoring the new conformations both as a complex and as unbound components without retaining the constraint potential,

$$I_{sc}^{\text{constr}} = S_{\text{bound}}^{\text{constr}} - S_{\text{unbound}}^{\text{constr}}$$

where the superscript indicates that the constrained conformations are used.

## Experimental Methods

### *Vector Construction*

The pMB1 origin of replication in the dual aminoacyl tRNA synthetase/tRNA expression vector, pRST.11B (Hughes and Ellington, 2010), was replaced by the p15 origin of replication from vector pACYC184 by amplification of the p15 origin from pACYC184 using primers p15AA.1, and p15A0.2 (**Supplementary Material**). The pRST.11B vector was amplified with primers VSP.2 and VSP.3 to generate a ~6kbp fragment that lacks the original origin of replication and is flanked by the unique SpeI and XmaI restriction endonuclease sites. The 1.2kbp PCR product containing the p15A origin and the 6kbp vector fragment was digested with SpeI and XmaI and ligated together to yield vector pRST.11C. The Nap1 mutant *Methanococcus jannaschii* tyrosyl amber suppressor tRNA (Guo et al., 2009) was assembled via PCR from the four overlapping oligonucleotides Nap.1, Nap.2, Nap.3 and Nap.4 (**Supplemental Material**). The assembled tRNA gene was digested with KpnI and BsrGI restriction enzymes and ligated into a similarly digested pRST.11C vector to yield vector pRST.11C-Nap1. A redundant XbaI site in pRST.11C-NapI was removed by quick change mutagenesis using primers, Qcxbaprstc.1 and Qcxbaprstc.2. The gene for the evolved L-DOPA utilizing aminoacyl tRNA synthetase (Alfonta et al., 2003) was assembled from overlapping oligonucleotides in-house using automated protein fabrication gene assembly process (Cox et al., 2007). The assembled gene was digested with

XbaI and XhoI and ligated into a similarly digested pRST.11C vector to yield the L-DOPA incorporating tRNA synthetase/tRNA vector, pDopa.

The pAK400-M18 scFv antibody and pAK400-pAD4 expression vectors (Leysath et al., 2009) were obtained from George Georgiou's group at the University of Texas at Austin. Amber (TAG) codons were introduced into the coding sequence of the M18 antibody via quick change mutagenesis or Gibson Assembly PCR.

#### *Expression and purification of M18 variants with NCAs.*

The M18 antibody variants were expressed using a condensed culture labeling method (Liu et al., 2010) in the presence (or absence) of supplemented L-DOPA. Briefly, the M18 antibody and variants were expressed by inoculating 900 mLs of 2xYT media containing 35 µg/mL chloramphenicol and 100 µg/mL ampicillin with 1 mL of a saturated overnight culture. Expression cultures were grown at 37°C to OD<sub>600</sub> ~0.8. Cultures were centrifuged at 3500g for 10 minutes, and resuspended in 100 mL 2xYT containing 5 mM DTT, 1.5 mM L-DOPA, and 1.5 mM IPTG. Condensed cultures were grown at 26°C for 12 hours. The PAD4 antigen was grown in Terrific Broth, induced at OD<sub>600</sub> = 1.0 with 1 mM IPTG, and allowed to grow at 30°C for 12 hours.

Expression cultures were centrifuged at 3500g for 15 minutes, and resuspended in PBS with 1mg/mL lysozyme and 0.25U/mL benzonase, and incubated on ice for 30 minutes. Cells were then sonicated for 4 minutes to further lyse the cells. Lysates were centrifuged at 35,000g for 45 minutes, after which the liquid fraction was poured over a 1.5 mL Ni-NTA agarose column. The resin was washed with 45 mL wash buffer (60 mM imidazole, 200 mM NaCl, 50 mM Phosphate at pH 8), and the protein was eluted with elution buffer (400 mM imidazole, 200 mM NaCl, 50 mM phosphate at pH 8). The Proteins were concentrated using an Amiconcellulose based centrifugal concentrators. Concentrations of the proteins were determined using Abs<sub>280</sub> measurements.

#### *Crosslinking Assay*

81 pmol (2.7 µM) PAD4 and 81 pmol (2.7 µM) of the M18 scFv variant were mixed in assay buffer (200 mM NaCl, 50 mM Phosphate at pH 8.5). Crosslinking was accomplished by adding



sodium periodate to 3.3 mM. Samples were incubated at room temperature for 30 minutes. Periodate was quenched with addition of DTT to 100 mM and 5x SDS loading dye. Samples were denatured by heating to 98 °C for ten minutes.

Samples were then run on Novex 4-12% Bis-tris SDS gels using MES-SDS running buffer at 200 V for 35 minutes. Proteins were transferred to nitrocellulose at 25 V for 1 hour, using Invitrogen XCell II Blot Module. Nitrocellulose was blocked at room temperature for 30 minutes using Superblock in PBS (Thermo) (for  $\alpha$ -His antibody) or Superblock in TBS (Thermo) (for  $\alpha$ -Flag antibody). Nitrocellulose was rinsed three times with PBS or TBS. Nitrocellulose was immersed in 50 mL of PBS or TBS with 10  $\mu$ L of  $\alpha$ -His or  $\alpha$ -Flag for 1 hour at room temperature. Nitrocellulose was rinsed with PBS or TBS three times, and resolved with Promega Western Blue Stabilized AP substrate for  $\alpha$ -His-AP conjugate, or with Promega ECL western blotting substrate for luminescent detection using the  $\alpha$ -Flag-HRP conjugate.

M18 variants were resolved at ~30 kDa, PAD4 resolves at 14 kDa. Crosslinked product appears at ~35 kDa. Luminescent western bands were quantified using ImageJ.

## RESULTS

We wished to test whether we could use computational design for the precise placement of crosslinks between antibodies and antigens. Our general plan was to use simple criteria and the Rosetta suite of software (Kaufmann et al., 2010; Leaver-Fay et al., 2011) to identify where we might place the NCAA L-DOPA, which upon activation to form the ortho-quinone in the presence of periodate (Burdine et al., 2004; Liu et al., 2006) can crosslink to a variety of residues, in particular the histidine imidazole and the  $\epsilon$ -amine of lysine. Thus, we first used Rosetta to build structural models of the antibody-antigen complex with candidate L-DOPA mutations, using energy parameters and rotamer libraries recently developed for NCAAs (Renfrew et al., 2012). For each L-DOPA position on the antibody, we evaluated various biophysical measurements that might affect crosslinking to an adjacent histidine or lysine on the bound antigen. The incorporation of L-DOPA into proteins was accomplished using an evolved tRNA synthetase from *Methanococcus jannaschii* that when expressed in *E. coli* can specifically

charge a suppressor tRNA with L-DOPA (Alfonta et al., 2003) . In the following sections we detail each of these steps.

## **Biophysical Criteria for Rational Design**

### *Initial choice of residues for L-DOPA substitution*

The X-ray structure of the M18-PAD4 complex, solved at 3.8 Å resolution, reveals a binding interface of  $\sim 1700 \text{ \AA}^2$ , with roughly equal contributions from the M18 light and heavy chains. All the antibody CDR loops except L3 contact the antigen, and CDR loops H3 and L2 appear to have the strongest interaction with PAD4 (Leysath et al., 2009). His and Lys residues are the only nucleophiles present on the antigen surface, however only Lys residues are near the epitope (**Figure 1**). To identify candidate positions, we selected all pairs of antigen nucleophilic residues and antibody residues that have heavy atoms within 10 Å. Thus we considered three nucleophilic lysines (PAD4 residues 684, 679 and 653) as crosslinking targets for sixteen possible L-DOPA locations on the antibody (**thirteen** on the heavy chain and **three** on the light chain).

Several biophysical parameters can be suggested as criteria for the evaluation and prediction of sites for L-DOPA insertion. First, the distance between the L-DOPA mutant and the reactive antigen nucleophilic residue should be close enough for them to react. Second, L-DOPA should be introduced to the antibody position where it is exposed to the solvent, so that periodate anions can access the reaction pair and trigger the reaction. Third, the original binding of M18/PAD4 should not be affected by the introduction of L-DOPA, meaning the interface must be compatible with it, or the change of interface energy should be moderate. Each of these criteria will be examined in turn as the basis for design.

### *Crosslinking Distance between L-DOPA and nucleophile*

The most obvious structural criterion for L-DOPA crosslinking is that the residue must be close enough to a nucleophile to crosslink. To assess this criterion, we took those L-DOPA atoms that were already within 10 Å of a lysine, as measured from  $C_\beta$  to  $C_\beta$ , and further calculated the

distance between C<sub>2</sub>, C<sub>5</sub>, and C<sub>6</sub> on the ring and the lysine N<sub>ε</sub> atom. In preliminary calculations, Rosetta tended to extend both L-DOPA and lysine residues into solution, and thus the calculated distances would exclude the possibility of crosslinking. To bring the reactive atoms closer, we optimized side-chain conformations using an energy function augmented with a constraint potential (see Methods). The crosslinking distance thus obtained,  $d_{XL}$ , varied from 2.9 to 9.8 Å (Table 1). For comparison, we also analyzed one L-DOPA substitution at a position far from the interface (L\_D17,  $d_{XL} = 32.1$  Å).

### *Solvent Accessibility of L-DOPA*

A second hypothesis is that a good crosslinking site needs access to the solvent so that the L-DOPA can be easily oxidized by periodate anions. In the presence of periodate, the hydroxyl groups on the phenolic ring of L-DOPA on M18 are oxidized to form the quinone intermediate, which activates the ring for potential nucleophilic attack by a neighboring lysine nucleophile on PAD4. We chose two methods to quantify the degree of solvent exposure. First, a simple metric was to count of the number of neighboring antibody and antigen atoms within 3 Å of the L-DOPA side chain. Mutants predicted to have fewer neighboring atoms were expected to have better chances for effective crosslinking. Second, the solvent-accessible surface area was calculated using a 1.4 Å probe. Both calculations were performed on the structure of antibody-antigen complex in which the L-DOPA and neighboring residue side chains were optimized by packing based on a rotamer library and the standard Rosetta “fixbb” protocol (see Computational Methods) to find an energetically favorable conformation. We ultimately sought to test whether there was a minimal solvent accessibility or whether crosslinking would improve as solvent accessibility increased (Table 1).

### *Mutant Interface Compatibility*

While it is difficult to computationally determine the impact of L-DOPA substitution on antigen binding, we tested whether Rosetta’s estimated energetics of the substituted interface would be helpful. Two energy measures were tested. First, we hypothesized that a successful crosslink would require that the introduced L-DOPA would not disrupt the favorable energy of interaction between the antigen and antibody. We therefore calculated the interface score,  $I_{sc}$ , a measure

used in docking calculations to approximate a binding energy (see Methods) (Chaudhury and Gray, 2008; Chaudhury et al., 2011). Data from docking 116 complexes (Chaudhury et al., 2011) indicate that most protein-protein complexes have negative interface scores in the range of -4 to -12 REU, and comparisons to experimental alanine mutants suggest that *changes* in interface scores above +1 REU are likely to characterize mutations that will significantly reduce binding affinity (Kortemme and Baker, 2002). One complication is that NCAs necessitate use of a Rosetta force field variation. In the case of the M18-PAD4 complex, the predicted interface score was -21.6 with the Rosetta's (standard) *score12* and +2.7 with the NCAA score function, *mm\_std*. The *mm\_std* interface scores of the L-DOPA substituted variants ranged from -1.4 to 4.8 REU, with most interface scores being slightly positive, indicating that the L-DOPA mutated M18-PAD4 complex may be slightly disfavored.

A second energetic hypothesis is that an activated conformation with proximity between L-DOPA and the target lysine should be energetically accessible. Thus, we evaluated the interface score for the structure we generated with the crosslinking constraint that brought the reacting atoms together ( $I_{sc}^{constr}$ ; see Methods). **Table 1** shows that these values vary from -0.6 to over 150 kcal/mol, showing that some cases exhibit significant energetic perturbations (e.g. H\_D56) and even atomic clashes (H\_R99 or H\_D54) when the L-DOPA/Lys pair is constrained.

## Crosslinking Studies

We chose to test candidate mutations based on their predicted placement within the protein-protein interface. We anticipated that the distance between the L-DOPA and a lysine nucleophile was likely to be the most critical parameter for successful crosslinking. Therefore, we made the closest eight substitutions within M18 (**Table 1**) (crosslinking distance of between 2.9 and 8.3 Å), excluding only H\_R99 because the high constrained interface energy ( $I_{sc}^{constr}$ ) suggested that it would bind antigen poorly. As controls for the crosslinking distance, two additional mutants were made by placing L-DOPA at two locations that were predicted to be unlikely to crosslink: H\_I51 (9.8 Å, solvent excluded) and L\_D17 (32.1 Å, solvent exposed).

The recombinant M18 antibodies were expressed in the periplasm of *E. coli* as a 28.9kDa single-chain variable fragment (scFv) with a N-terminal FLAG epitope tag and a C-terminal His tag to

enable purification via immobilized metal affinity chromatography (IMAC). This tag also allows ready detection of the antibody via Western blot analysis. Similarly, the PAD4 antigen was expressed as a 17.5kDa protein with a C-terminal His tag to enable purification of the antigen via IMAC. L-DOPA was introduced site-specifically into the chosen sites in the CDRs of M18 by changing the codon for the naturally occurring amino acid at these positions to an Amber (TAG) stop codon. The M18 amber variants were then co-expressed with an orthogonal tRNA suppressor and tRNA synthetase pair that had previously been evolved to be specific for L-DOPA (Alfonta et al., 2003). Each M18 variant was incubated with PAD4, and crosslinking was catalyzed by the addition of sodium periodate. Crosslinking between predicted L-DOPA-containing M18 variants and PAD4 was probed and quantitated using a Western blot assay to the FLAG epitope tag (**Figure 2**). A covalent crosslink between antibody and antigen was apparent by a shift in the molecular mass of the antibody. The efficiency of crosslinking was determined by the percentage of antibody that underwent a mass shift. Each variant was tested a minimum of two times. The distal crosslinking controls (H\_I51 and L\_D17) showed no appreciable crosslinking (< 3%), and eight of the substitutions exhibited crosslinking values ranging from 5% to 55% (**Table 1**).

### Correlations between Rosetta criteria and crosslinking

We can inspect the biophysical criteria and determine whether and how they impact crosslinking efficiency. **Figure 3** shows the plots of each of the five biophysical criteria described above versus the experimentally observed crosslinking extent. Crosslinking efficiency dropped off appreciably (< 3%) at distances greater than 7 Å [*e.g.* variants H\_G55 (8.3Å), H\_I51 (9.8Å), and L\_D17 (32.1 Å)] relative to L-DOPA substitutions that were predicted to lie closer to lysines (**Figure 3a**). There was no correlation between solvent accessibility measures and crosslinking efficiency when all data are considered (**Figure 3b-c**). However, when the points representing mutants with long-distance crosslinks are excluded (black in **Figure 3b-c**), both the number of residues that surround an L-DOPA substitution and the solvent accessible surface area (SASA) show relationships to crosslinking efficiency (**Figure 3c**). In particular, variants H\_Y52 and H\_W33, with SASA of only 25.1 and 14.1 Å<sup>2</sup> respectively, have the lowest crosslinking efficiencies (19.7% and 13.9%, respectively) amongst variants with  $d_{XL}$  under 7 Å.

The energetic parameters do not provide clear trends for the analysis (or prediction) of crosslinking efficiency even after filtering for distance and solvent accessibility (**Figure 3d-e**). However, it is notable that the two most efficient crosslinking sites (H\_S31 and H\_G53) have low energies ( $I_{sc}$  and  $I_{sc}^{constr}$  of 1 REU or less in both cases). At best, we can say that by also choosing positions that pass the energy filter, we may identify variants that exhibit efficient crosslinking. Overall, these plots emphasize the key importance of distance, followed by solvent accessibility, in identifying positions that may be capable of significant crosslinking.

### **Interface structure of a successful crosslinking case**

**Figure 4** shows a detail of the structural model of the best crosslinking case with the L-DOPA mutant at position H\_S31 targeting the Lys at position 679. This mutant exhibits 52% crosslinking (standard deviation 12%). The crosslinking distance is a moderate 5.7 Å, and the SASA of 132 Å<sup>2</sup> shows good solvent accessibility. The calculated interface energies of  $I_{sc} = 0.7$  REU and  $I_{sc}^{constr} = -0.6$  REU suggest that both the unconstrained and constrained conformations are energetically favorable in the context of the bound antibody-antigen complex. As seen in the figure, the aromatic carbon atoms C<sub>2</sub>, C<sub>5</sub>, and C<sub>6</sub> are all positioned for nucleophilic attack by the amine.

## **DISCUSSION**

The insertion of new, functional amino acids into proteins, even proteins of known structure, remains an enterprise that is fraught with uncertainty. While there have been several successful demonstrations of the rational placement of crosslinking amino acids into proteins (Alfonta et al., 2003; Forne et al., 2012; Sato et al., 2011; Umeda et al., 2009; Umeda et al., 2010), to date there has not been a thorough analysis of how a detailed structural analysis might be used to guide the placement of amino acids that would crosslink efficiently. **Currently, placement of NCAs into proteins is based on intuition and no computational tools exist to guide the design of crosslinks.** As a first step towards the development of tools for engineering crosslinks, we sought to develop a rational, rules-based approach for predicting the placement of a crosslinking NCA within a protein-protein complex. The placement of L-DOPA was correlated with its ability to crosslink,

and the biophysical parameters that most impacted successful crosslinking were compared. Based on computational predictions, a number of high efficiency crosslinking sites were recovered. The best engineered antibody could crosslink with very high efficiency (55%) with little to no observable side reactions. Additional optimizations of crosslinking by varying buffer conditions and periodate concentrations may lead to further enhancements in crosslinking efficiency.

Our results are especially significant because typical crosslinking reactions for rationally placed variants have been reported to suffer from a lack of specificity, low yield, and uncharacterized side reactions (Fancy and Kodadek, 1999; Fancy et al., 1996; Sinz, 2006). Additionally, several commercially available ‘random’ crosslinkers have been reported to be inefficient and barely detectable by MS analysis (Dihazi and Sinz, 2003). In our study, we found that adjacency is an excellent primary criterion for success, and additionally, that greater solvent accessibility increases crosslinking efficiency for proximal residues. Although energetic calculations were noisy, the best two positions for crosslinking L-DOPA moieties showed low interface energies, both when free and when constrained to be near the nucleophile.

Direct comparison of the crosslinking efficiencies obtained herein to previously published results proved difficult because crosslinking efficiencies have generally not been evaluated quantitatively, but rather just examined qualitatively by gel mobility shifts (Burdine et al., 2004; Umeda et al., 2009; Umeda et al., 2010). Our results therefore likely represent the first quantitative study of L-DOPA crosslinking.

That said, the utility of the computational methods employed can perhaps be discerned by comparing them with intuition. In the case of antibody M18 and PAD, positions H\_D54, H\_D56, or H\_Y52 on the antibody would likely have been chosen as the best locations for L-DOPA, since they lie close to a lysine in the complex. While L-DOPA insertions at all three of these positions did show crosslinking (19.7-36.2%), the most efficient crosslinking (40.3% and 51.9%) instead occurred at the computer-chosen positions H\_S31 and H\_R99 that scored well by all three biophysical criteria: distance, solvent accessibility, and energetics.

As this study relied on a single antibody-antigen target with a known crystal structure of the complex the conclusions should be generalized with caution. Nonetheless, a strategy for identifying high-probability antibody-antigen crosslinking positions might consist of a three-part filter that requires: (i) crosslinking distance under 7 Å; (ii) sufficient solvent-accessibility (SASA over 90 Å<sup>2</sup> or fewer than ten neighboring atoms); and (iii) compatible interface energies ( $I_{sc}$  under 2.5 REU and  $I_{sc}^{constr}$  under 10 REU). Application of these filters in this case led to the quick and accurate prediction of a small set of substitutions from which generally excellent crosslinking antibodies were derived. This differs from previous attempts where iterations of insertion and crosslinking conditions had to be considered in order to identify functional crosslinks. Further studies will be needed to test the utility of these approaches with antibody homology models (e.g. (Marcatili et al., 2008),(Pantazes and Maranas, 2010),(Sivasubramanian et al., 2009)) or docked structures (e.g. (Sircar and Gray, 2010)).

## ACKNOWLEDGEMENT

This project was supported by the DARPA Antibody Technology Program (HR-0011-10-1-0052). The authors thank Dr. P. Douglas Renfrew for his help implementing non-canonical amino acids in Rosetta, and we thank George Georgiou, Brian Kuhlman, and Bryan Der for helpful discussions.

## REFERENCES

- A. Caravella, J., Wang, D., M. Glaser, S., Lugovskoy, A., 2010. Structure-Guided Design of Antibodies. *Curr. Comput. Aided. Drug. Des.* 6, 128-138.
- Ai, H.W., Shen, W., Sagi, A., Chen, P.R., Schultz, P.G., 2011. Probing protein-protein interactions with a genetically encoded photo-crosslinking amino acid. *Chembiochem* 12, 1854-1857.
- Alfonta, L., Zhang, Z., Uryu, S., Loo, J.A., Schultz, P.G., 2003. Site-Specific Incorporation of a Redox-Active Amino Acid into Proteins. *J. Am. Chem. Soc.* 125, 14662-14663.



- Ayyadurai, N., Prabhu, N.S., Deepankumar, K., Jang, Y.J., Chitrapriya, N., Song, E., Lee, N., Kim, S.K., Kim, B.G., Soundrarajan, N., Lee, S., Cha, H.J., Budisa, N., Yun, H., 2011. Bioconjugation of L-3,4-dihydroxyphenylalanine containing protein with a polysaccharide. *Bioconjug. Chem.* 22, 551-555.
- Barderas, R., Desmet, J., Timmerman, P., Melen, R., Casal, J.I., 2008. Affinity maturation of antibodies assisted by in silico modeling. *Proc. Natl. Acad. Sci. USA* 105, 9029-9034.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The Protein Data Bank. *Nucl. Acids Res.* 28, 235-242.
- Burdine, L., Gillette, T.G., Lin, H.-J., Kodadek, T., 2004. Periodate-Triggered Cross-Linking of DOPA-Containing Peptide-Protein Complexes. *J. Am. Chem. Soc.* 126, 11442-11443.
- Chaudhury, S., Gray, J.J., 2008. Conformer Selection and Induced Fit in Flexible Backbone Protein-Protein Docking Using Computational and NMR Ensembles. *J. Mol. Bio.* 381, 1068-1087.
- Chaudhury, S., Berrondo, M., Weitzner, B.D., Muthu, P., Bergman, H., Gray, J.J., 2011. Benchmarking and Analysis of Protein Docking Performance in Rosetta v3.2. *PLoS ONE* 6, e22477.
- Chennamsetty, N., Voynov, V., Kayser, V., Helk, B., Trout, B.L., 2009. Design of therapeutic proteins with enhanced stability. *Proc. Natl. Acad. Sci. USA* 106, 11937-11942.
- Chin, J.W., Martin, A.B., King, D.S., Wang, L., Schultz, P.G., 2002a. Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. *Proc Natl Acad Sci U S A* 99, 11020-11024.
- Chin, J.W., Santoro, S.W., Martin, A.B., King, D.S., Wang, L., Schultz, P.G., 2002b. Addition of p-azido-L-phenylalanine to the genetic code of *Escherichia coli*. *J Am Chem Soc* 124, 9026-9027.

- Cho, H., Daniel, T., Buechler, Y.J., Litzinger, D.C., Maio, Z., Putnam, A.-M.H., Kraynov, V.S., Sim, B.-C., Bussell, S., Javahishvili, T., Kaphle, S., Viramontes, G., Ong, M., Chu, S., GC, B., Lieu, R., Knudsen, N., Castiglioni, P., Norman, T.C., Axelrod, D.W., Hoffman, A.R., Schultz, P.G., DiMarchi, R.D., Kimmel, B.E., 2011. Optimized clinical performance of growth hormone with an expanded genetic code. *Proc. Natl. Acad. Sci. USA*.
- Clark, L.A., Boriack-Sjodin, P.A., Eldredge, J., Fitch, C., Friedman, B., Hanf, K.J.M., Jarpe, M., Liparoto, S.F., Li, Y., Lugovskoy, A., Miller, S., Rushe, M., Sherman, W., Simon, K., Van Vlijmen, H., 2006. Affinity enhancement of an in vivo matured therapeutic antibody using structure-based computational design. *Protein Sci.* 15, 949-960.
- Cox, J.C., Lape, J., Sayed, M.A., Hellinga, H.W., 2007. Protein fabrication automation. *Protein Sci.* 16, 379-390.
- Dihazi, G.H., Sinz, A., 2003. Mapping low-resolution three-dimensional protein structures using chemical cross-linking and Fourier transform ion-cyclotron resonance mass spectrometry. *Rapid Commun Mass Spectrom* 17, 2005-2014.
- Fancy, D.A., Kodadek, T., 1999. Chemistry for the analysis of protein-protein interactions: rapid and efficient cross-linking triggered by long wavelength light. *Proc Natl Acad Sci U S A* 96, 6020-6024.
- Fancy, D.A., Melcher, K., Johnston, S.A., Kodadek, T., 1996. New chemistry for the study of multiprotein complexes: the six-histidine tag as a receptor for a protein crosslinking reagent. *Chem Biol* 3, 551-559.
- Farady, C.J., Sellers, B.D., Jacobson, M.P., Craik, C.S., 2009. Improving the species cross-reactivity of an antibody using computational design. *Bioorg. Med. Chem. Lett.* 19, 3744-3747.

- Forne, I., Ludwigsen, J., Imhof, A., Becker, P.B., Mueller-Planitz, F., 2012. Probing the conformation of the ISWI ATPase domain with genetically encoded photoreactive crosslinkers and mass spectrometry. *Mol. Cell Proteomics* 11, M111 012088.
- Guo, J., Melancon, C.E., 3rd, Lee, H.S., Groff, D., Schultz, P.G., 2009. Evolution of amber suppressor tRNAs for efficient bacterial production of proteins containing nonnatural amino acids. *Angew. Chem. Int. Ed. Engl.* 48, 9148-9151.
- Harvey, B.R., Georgiou, G., Hayhurst, A., Jeong, K.J., Iverson, B.L., Rogers, G.K., 2004. Anchored periplasmic expression, a versatile technology for the isolation of high-affinity antibodies from *Escherichia coli*-expressed libraries. *Proc. Natl. Acad. Sci. USA* 101, 9193-9198.
- Hughes, R.A., Ellington, A.D., 2010. Rational design of an orthogonal tryptophanyl nonsense suppressor tRNA. *Nucl. Acids Res.* 38, 6813-6830.
- Jones, D.H., Cellitti, S.E., Hao, X., Zhang, Q., Jahnz, M., Summerer, D., Schultz, P.G., Uno, T., Geierstanger, B.H., 2010. Site-specific labeling of proteins with NMR-active unnatural amino acids. *J. Biomol. NMR* 46, 89-100.
- Kaufmann, K.W., Lemmon, G.H., DeLuca, S.L., Sheehan, J.H., Meiler, J., 2010. Practically Useful: What the Rosetta Protein Modeling Suite Can Do for You. *Biochemistry* 49, 2987-2998.
- Kim, J., Yoon, M.-Y., 2010. Recent advances in rapid and ultrasensitive biosensors for infectious agents: lesson from *Bacillus anthracis* diagnostic sensors. *Analyst* 135, 1182-1190.
- Kortemme, T., Baker, D., 2002. A simple physical model for binding energy hot spots in protein-protein complexes. *Proc. Natl. Acad. Sci. USA* 99, 14116-14121.
- Kuroda, D., Shirai, H., Jacobson, M.P., Nakamura, H., 2012. Computer-aided antibody design. *Protein Eng. Des. Sel.* 25, 507-522.

- Le Grand, S.M., Merz, K.M., 1993. Rapid approximation to molecular surface area via the use of Boolean logic and look-up tables. *J. Comp. Chem.* 14, 349-352.
- Leaver-Fay, A., Tyka, M., Lewis, S.M., Lange, O.F., Thompson, J., Jacak, R., Kaufman, K.W., Renfrew, P.D., Smith, C.A., Sheffler, W., Davis, I.W., Cooper, S., Treuille, A., Mandell, D.J., Richter, F., Ban, Y.-E.A., Fleishman, S.J., Corn, J.E., Kim, D.E., Lyskov, S., Berrondo, M., Mentzer, S., Popovič, Z., Havranek, J.J., Karanicolas, J., Das, R., Meiler, J., Kortemme, T., Gray, J.J., Kuhlman, B., Baker, D., Bradley, P., 2011. Chapter nineteen - Rosetta3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules, p. 545-574, in: L. J. Michael and B. Ludwig, Eds.), *Methods Enzymol.*, Academic Press.
- Leysath, C.E., Monzingo, A.F., Maynard, J.A., Barnett, J., Georgiou, G., Iverson, B.L., Robertus, J.D., 2009. Crystal Structure of the Engineered Neutralizing Antibody M18 Complexed to Domain 4 of the Anthrax Protective Antigen. *J. Mol. Bio.* 387, 680-693.
- Lippow, S.M., Tidor, B., 2007. Progress in computational protein design. *Curr. Opin. Biotechnol.* 18, 305-311.
- Lippow, S.M., Wittrup, K.D., Tidor, B., 2007. Computational design of antibody-affinity improvement beyond in vivo maturation. *Nat. Biotech.* 25, 1171-1176.
- Little, S.F., Leppla, S.H., Cora, E., 1988. Production and characterization of monoclonal antibodies to the protective antigen component of *Bacillus anthracis* toxin. *Infect. Immun.* 56, 1807-1813.
- Liu, B., Burdine, L., Kodadek, T., 2006. Chemistry of Periodate-Mediated Cross-Linking of 3,4-Dihydroxyphenylalanine-Containing Molecules to Proteins. *J. Am. Chem. Soc.* 128, 15228-15235.

- Liu, C.C., Choe, H., Farzan, M., Smider, V.V., Schultz, P.G., 2009. Mutagenesis and evolution of sulfated antibodies using an expanded genetic code. *Biochemistry* 48, 8891-8898.
- Liu, J., Castaneda, C.A., Wilkins, B.J., Fushman, D., Cropp, T.A., 2010. Condensed *E. coli* cultures for highly efficient production of proteins containing unnatural amino acids. *Bioorg. Med. Chem. Lett.* 20, 5613-5616.
- Mandell, D.J., Kortemme, T., 2009. Computer-aided design of functional protein interactions. *Nat. Chem. Biol.* 5, 797-807.
- Marcatili, P., Rosi, A., Tramontano, A., 2008. PIGS: automatic prediction of antibody structures. *Bioinformatics* 24, 1953-1954.
- Midelfort, K.S., Hernandez, H.H., Lippow, S.M., Tidor, B., Drennan, C.L., Wittrup, K.D., 2004. Substantial energetic improvement with minimal structural perturbation in a high affinity mutant antibody. *J Mol Biol* 343, 685-701.
- Miklos, A.E., Kluwe, C., Der, B.S., Pai, S., Sircar, A., Hughes, R.A., Berrondo, M., Xu, J., Codrea, V., Buckley, P.E., Calm, A.M., Welsh, H.S., Warner, C.R., Zacharko, M.A., Carney, J.P., Gray, J.J., Georgiou, G., Kuhlman, B., Ellington, A.D., 2012. Structure-Based Design of Supercharged, Highly Thermoresistant Antibodies. *Chem. Biol.* 19, 449-455.
- Moayeri, M., Leppla, S.H., 2004. The roles of anthrax toxin in pathogenesis. *Curr. Opin. Microbiol.* 7, 19-24.
- Pantazes, R.J., Maranas, C.D., 2010. OptCDR: a general computational method for the design of antibody complementarity determining regions for targeted epitope binding. *Protein Eng. Des. Sel.* 23, 849-858.

- Pimenova, T., Nazabal, A., Roschitzki, B., Seebacher, J., Rinner, O., Zenobi, R., 2008. Epitope mapping on bovine prion protein using chemical cross-linking and mass spectrometry. *J. Mass Spectrom.* 43, 185-195.
- Renfrew, P.D., Choi, E.J., Bonneau, R., Kuhlman, B., 2012. Incorporation of Noncanonical Amino Acids into Rosetta and Use in Computational Protein-Peptide Interface Design. *PLoS ONE* 7, e32637.
- Sato, S., Mimasu, S., Sato, A., Hino, N., Sakamoto, K., Umehara, T., Yokoyama, S., 2011. Crystallographic study of a site-specifically cross-linked protein complex with a genetically incorporated photoreactive amino acid. *Biochemistry* 50, 250-257.
- Simons, K.T., Ruczinski, I., Kooperberg, C., Fox, B.A., Bystroff, C., Baker, D., 1999. Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. *Proteins: Struct., Funct. Genet.* 34, 82-95.
- Sinz, A., 2006. Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrom Rev* 25, 663-682.
- Sircar, A., Gray, J.J., 2010. SnugDock: Paratope Structural Optimization during Antibody-Antigen Docking Compensates for Errors in Antibody Homology Models. *PLoS Comput. Biol.* 6, e1000644.
- Sivasubramanian, A., Sircar, A., Chaudhury, S., Gray, J.J., 2009. Toward high-resolution homology modeling of antibody Fv regions and application to antibody-antigen docking. *Proteins: Structure, Function, and Bioinformatics* 74, 497-514.
- Tsao, M.L., Summerer, D., Ryu, Y., Schultz, P.G., 2006. The genetic incorporation of a distance probe into proteins in *Escherichia coli*. *J. Am. Chem. Soc.* 128, 4572-4573.

- Umeda, A., Thibodeaux, G.N., Zhu, J., Lee, Y., Zhang, Z.J., 2009. Site-specific protein cross-linking with genetically incorporated 3,4-dihydroxy-L-phenylalanine. *Chembiochem.* 10, 1302-1304.
- Umeda, A., Thibodeaux, G.N., Moncivais, K., Jiang, F., Zhang, Z.J., 2010. A versatile approach to transform low-affinity peptides into protein probes with cotranslationally expressed chemical cross-linker. *Anal. Biochem.* 405, 82-88.
- Wang, L., Xie, J., Schultz, P.G., 2006. EXPANDING THE GENETIC CODE. *Annu. Rev. Biophys. Biomol. Struct.* 35, 225-249.
- Xie, J., Schultz, P.G., 2006. A chemical toolkit for proteins - an expanded genetic code. *Nat. Rev. Mol. Cell Biol.* 7, 775-782.
- Young, J.A.T., Collier, R.J., 2007. Anthrax Toxin: Receptor Binding, Internalization, Pore Formation, and Translocation. *Annu. Rev. Biochem.* 76, 243-265.
- Zhang, Z., Wang, L., Brock, A., Schultz, P.G., 2002. The selective incorporation of alkenes into proteins in *Escherichia coli*. *Angew. Chem. Int. Ed. Engl.* 41, 2840-2842.

## FIGURE CAPTIONS

**Figure 1:** The M18/PAD4 antibody-antigen complex structure from PDB 3ETB highlighting the antigen's candidate crosslinking residues, Lys (cyan sticks) and His (yellow sticks).

**Figure 2:** A western blot of M18 and variants treated with periodate in the absence and presence of antigen PAD4.  $\alpha$ -Flag-HRP was used to detect a Flag-tag on the M18 variants. The shift in mass in the presence of PAD4 corresponds to the mass of PAD4 antigen (~14 kDa), signaling covalent crosslinking between antibody and antigen. Bands were quantified using ImageJ to determine crosslinking efficiencies, which were listed in Table 1. D17L is the only tested position on M18 light chain. The rest of variants were all on heavy chain.

**Figure 3:** Plots relating the biophysical criteria to experimental crosslinking efficiency. Error bars display the standard deviations obtained from multiple trials in experiments. (a). Effect of crosslinking distance. (b,c). Effect of solvent accessibility by (b) L-DOPA neighboring atoms or (c) solvent-accessible surface area. **Black circles represent positions with  $d_{XL}$  over 7 Å (H\_G55, H\_I51, and L\_D17).** (d,e). Effect of interface energetics by (d) interface energy or (e) crosslink fluctuation energy. **Again black circles represent positions with  $d_{XL}$  over 7 Å, and additionally red circles represent positions with solvent-accessible surface area under 60 Å (H\_Y52 and H\_W33).**

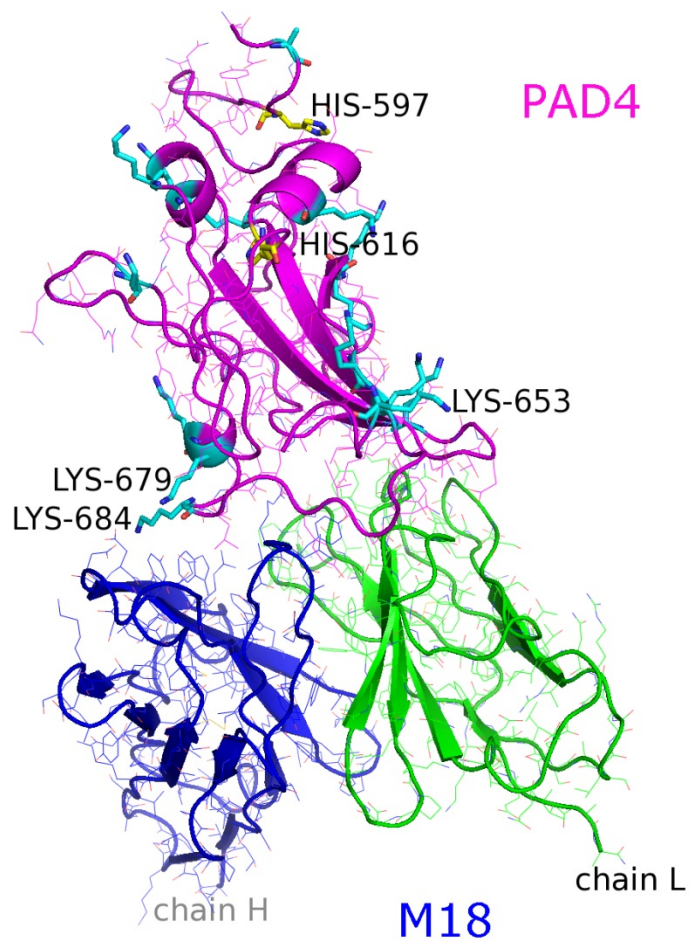
**Figure 4:** Structural detail of the successfully crosslinking L-DOPA mutant at position H\_S31 with its targeted Lys at position 679. Crosslinking distance of 5.7 Å, 132 Å<sup>2</sup> of solvent-accessible surface area, and calculated interface energies of  $I_{sc} = 0.7$  REU and  $I_{sc}^{constr} = -0.6$  REU. Aromatic carbon atoms C<sub>2</sub>, C<sub>5</sub>, and C<sub>6</sub> are all positioned for nucleophilic attack by the amine. The crosslinking distances for these three sites are 7.8 Å, 6.3 Å, and 5.7 Å, respectively.



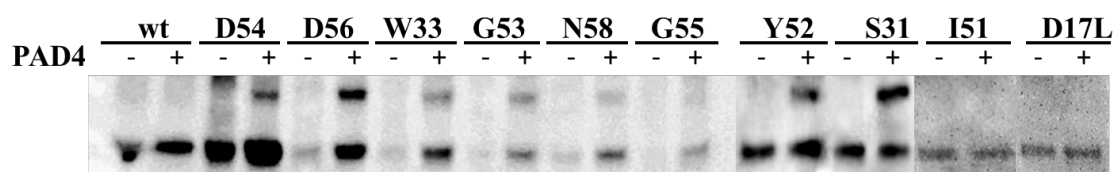
**Table 1. Candidate crosslink residue pairs, calculated biophysical properties, and experimental crosslinking of the M18/PAD4 interface with an L-DOPA mutation.** Antigen residue nucleophile residues paired with antibody positions for placement of L-DOPA. The calculated values of factors (accessible distance, number of neighbors, interface score) after L-DOPA mutation that can affect L-DOPA-Lys crosslinking. Mutants are sorted by accessible distance. Grey shading indicates poor conditions that may reduce or prevent crosslinking ( $d_{XL} > 7 \text{ \AA}$ , over ten neighbors,  $SASA < 90 \text{ \AA}^2$ ,  $I_{sc} > 2.5 \text{ REU}$ , or  $I_{sc}^{constr} > 10 \text{ REU}$ ). For the two cases where antibody positions have multiple candidate crosslink targets, only one is shown in the table (Position H\_G53's second target is J\_K684 at  $d_{XL} = 4.1 \text{ \AA}$  with  $I_{sc} = 1.0 \text{ REU}$  and  $I_{sc}^{constr} = 11.5 \text{ REU}$ , and position H\_S31's second target is J\_K684 at  $d_{XL} = 8.6 \text{ \AA}$  with  $I_{sc} = 0.7 \text{ REU}$  and  $I_{sc}^{constr} = 4.7 \text{ REU}$ .)

Antigen Residue	Antibody Position	Crosslinking		SASA ( $\text{\AA}^2$ )	Interface Score, $I_{sc}$ (REU)	Interface Score $I_{sc}^{constr}$ (REU)	Experimental Crosslinking
		Distance, $d_{XL}$ ( $\text{\AA}$ )	Number of Neighbors				
J_K684	H_D54	2.9	4	124.9	0.8	52.0	30.4% (15.8%)
J_K684	H_D56	3.0	2	120.3	1.1	28.9	36.2% (5.0%)
J_K684	H_Y52	3.0	14	25.1	1.3	15.6	19.7% (4.4%)
J_K684	H_W33	3.4	9	14.1	4.8	8.4	13.9% (8.4%)
J_K684	H_N58	5.5	6	111.3	3.7	6.1	24.4% (3.3%)
J_K679	H_S31	5.7	1	132.3	0.7	-0.6	51.9% (11.5%)
J_K679	H_G53	6.1	8	124.8	1.0	0.2	40.3% (13.8%)
J_K684	H_R99	6.6	8	28.2	0.2	151.1	not tested
J_K684	H_G55	8.3	1	161.6	2.8	1.8	5.1% (1.3%)
J_K653	L_N31	8.6	9	75.3	1.2	-1.1	not tested
J_K684	H_S57	8.8	2	122.7	2.7	12.7	not tested
J_K684	H_S32	8.9	18	42.9	1.3	12.2	not tested
J_K653	L_Y50	9.0	11	10.5	2.8	0.1	not tested
J_K684	H_G96	9.3	11	56.8	-1.4	11.5	not tested
J_K684	H_I51	9.8	22	0.5	2.8	13.4	2.7% (0.14%)
J_K653	L_D17	32.1	4	139.9	2.7	0.2	1.5% (0.49%)

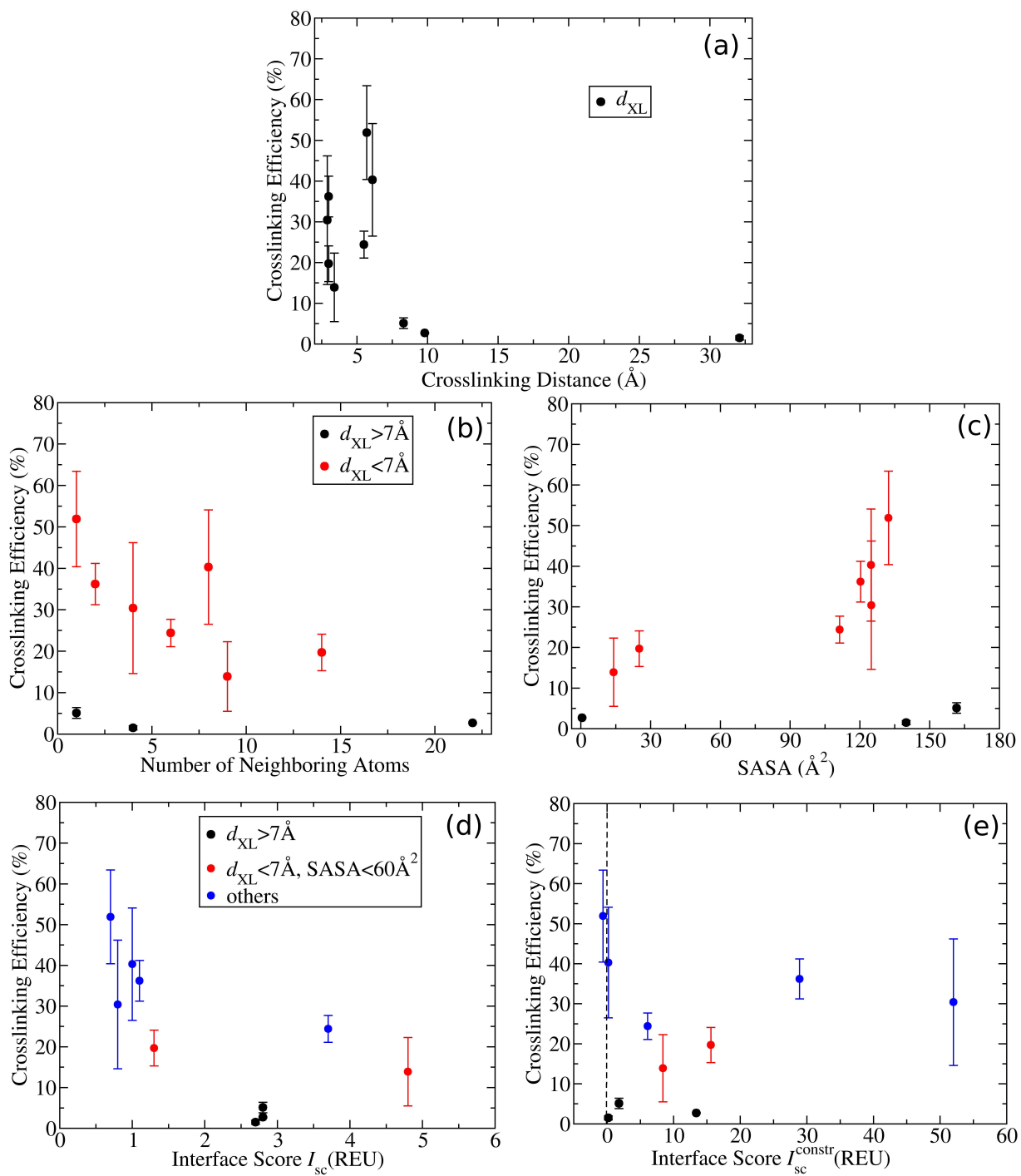
**Figure 1**



**Figure 2**



**Figure 3**



**Figure 4**

