Published in final edited form as:

Science. 2015 February 20; 347(6224): 863–867. doi:10.1126/science.aaa2424.

Crystal Structure of a Kinetically Persistent Transition State in a Computationally Designed Protein Bottle

Aaron D. Pearson^{#1}, Jeremy H. Mills^{#2}, Yifan Song², Fariborz Nasertorabi³, Gye Won Han³, David Baker^{2,4}, Raymond C. Stevens³, and Peter G. Schultz^{1,‡}

¹Department of Chemistry, and Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

²Department of Biochemistry, University of Washington, Seattle, WA, 98195, USA

³Department of Biological Sciences, University of Southern California, Los Angeles, CA 90089

⁴Howard Hughes Medical Institute (HHMI), University of Washington, Seattle, WA 98195, USA

[#] These authors contributed equally to this work.

Abstract

The fleeting lifetimes of the transition states (TSs) of chemical reactions make determination of their three-dimensional structures by diffraction methods a challenge. Herein we use packing interactions within the core of a protein to stabilize the planar TS for rotation around the central C-C bond of biphenyl so that it can be directly observed by x-ray crystallography. The computational protein design software Rosetta was used to design a pocket within threonyl-transfer RNA synthetase from the thermophile *Pyrococcus abyssi* that forms complementary van der Waals interactions with a planar biphenyl. This latter moiety was introduced biosynthetically as the side chain of the noncanonical amino acid *p*-biphenylalanine. Through iterative rounds of computational design and structural analysis we identified a protein in which the side chain of *p*-biphenylalanine is kinetically trapped in the energetically disfavored, coplanar conformation of the TS of the bond rotation reaction.

The direct observation of the transition state (TS) of a chemical reaction requires highly sensitive spectroscopic techniques with a temporal resolution on the order of 10^{-13} to 10^{-14} s. This milestone was made possible with the advent of ultrafast laser spectroscopy, which allowed the direct observation of the transient species formed as reactants cross an energy barrier to products (1, 2). However, determination of the three dimensional (3D) structure of a TS requires ultrafast electron or x-ray diffraction techniques with similar temporal resolution and sensitivity (3, 4). An alternative approach is to increase the lifetime of the transition state configuration by trapping it in a free energy (ΔG) well so that it can be directly observed by more conventional spectroscopic methods. In theory the well-packed interior of a protein could act as a “programmable” solvent to selectively stabilize a TS, and thereby make it kinetically persistent and directly observable by conventional x-ray diffraction methods. Indeed, enzymes catalyze reactions by virtue of their ability to

[‡]To whom correspondence should be addressed. schultz@scripps.edu.

selectively stabilize the rate-limiting TS relative to the ground-state reactants (5-9). Here we show the redesign of the interior of the protein threonyl-tRNA synthetase from the thermophile *Pyrococcus abyssi* (10) to create a microenvironment that has high complementarity to the planar TS configuration for rotation about the central C-C bond of biphenyl (11). Determination of the 3D structure of the designed protein at 2.05 Å resolution by x-ray crystallography revealed a planar biphenyl TS configuration stabilized by van der Waals interactions with side chains in the protein core.

Biphenyl bond rotation is a well-studied reaction, both experimentally and theoretically. In the gas phase the ground state of biphenyl is twisted with a dihedral angle of $\sim 45^\circ$. Rotation around the central C-C bond connecting the two phenyl rings in biphenyl is estimated to have an energy barrier of 5.8 ± 2.1 kJ/mol and 6.7 ± 2.1 kJ/mol around the 0° (planar) and 90° TSs, respectively, as determined experimentally by electron diffraction studies (Fig. 1 A) (12). The energy barriers estimated using Raman data (13, 14) and ultraviolet absorption spectroscopy (15) are in general agreement with the results from the diffraction studies. Theoretical calculations show that these barriers result from opposing steric and electronic effects of the two phenyl rings, but give a range of values depending on the method used (16). This simple reaction (which racemizes chiral biphenyls) provides an ideal system to pose the question whether side chain packing interactions in a folded protein can be exploited to make a transition state configuration kinetically persistent so that it can be directly observed by x-ray crystallography.

In order to introduce a biphenyl moiety into a protein core, we used the genetically encoded noncanonical amino acid, *p*-biphenylalanine (BiPhe). This amino acid can be introduced site-specifically into a protein in *Escherichia coli* in good yield in response to the amber nonsense codon, TAG, with an orthogonal amber suppressor tRNA/aminoacyl-tRNA synthetase (aaRS) pair. The aaRS was evolved from the *Methanococcus jannaschi* tyrosyl-tRNA synthetase to be selective for BiPhe, and not incorporate any of the 20 canonical amino acids (17, 18). The rotational barrier around the biphenyl C-C bond in BiPhe is not expected to be substantially different from biphenyl itself (19).

To identify an appropriate host protein in which to construct a site complementary site to the planar biphenyl TS geometry, we first curated a set of $\sim 2,300$ proteins of known structures from thermophilic organisms (a full description of the computational design process is found in the supporting information, (20)). Proteins with high thermostabilities have more negative

*G*s of folding and are in general more tolerant to mutations; both characteristics are likely to be beneficial when attempting to stabilize a TS configuration through packing interactions in a protein core (21, 22). RosettaMatch (23) was then used to identify residues in the core of each protein scaffold where BiPhe could be substituted without creating unfavorable steric interactions with the protein backbone in order to minimize changes to the global protein fold (Fig 1 B). In a second “matching” step, π -stacking interactions (24) between hydrophobic Trp, Phe, and Tyr residues in the native scaffolds and the substituted BiPhe were identified. Because the RosettaMatch calculations were carried out with a planar model of the BiPhe, these interactions should stabilize the biphenyl side chain in the desired planar TS conformation. To ensure that the substituted BiPhe side chain was sufficiently buried within a particular protein’s core, initial hits were first filtered on the basis of the change in

solvent-accessible surface area (SASA) (25) that occurs when BiPhe is removed from the substituted site. A SASA cutoff of 0.9 (meaning the BiPhe was > 90% buried) removed 35% of the initial matches from further consideration.

On the basis of the above analysis we chose threonyl-tRNA synthetase from the thermophile *Pyrococcus abyssi* as our initial mutagenesis scaffold. RosettaDesign (26) was then used to optimize the identities of residues surrounding the BiPhe (excluding residues already participating in π -stacking interactions) such that they packed tightly against the planar BiPhe side chain but did not create unfavorable hydrogen-bonding interactions (Fig. 1 B). Candidate designs were ranked by shape complementarity (SC) (27) between the designed residues and the BiPhe (Fig. 1 B). Ultimately, four designs (BIF_1–BIF_4) with high SC values were chosen for experimental characterization (Table 1). The computer-modeled proteins containing the BiPhe were reverse-translated and optimized for expression in *E. coli*.

To incorporate BiPhe into these four different protein scaffolds, we used the orthogonal amber suppressor tRNA/BiPheRS pair encoded on the dual plasmid expression system (pUltra) (28). One plasmid contained the tRNA/BiPheRS pair specific for BiPhe, and the other contained the synthetic mutant gene of interest fused to a C-terminal hexa-histidine purification tag. A TAG amber nonsense codon was introduced at the desired site to encode BiPhe (17). These plasmids were co-transformed into *E. coli* B121(DE3), and protein expression was carried out in the presence of 1mM BiPhe. Proteins were purified from cell lysate via Ni^{2+} -affinity chromatography followed by size-exclusion chromatography. Mass spectrometric analysis of BIF_1–BIF_4 indicated successful incorporation of BiPhe in all cases (Fig. S1), and SDS-PAGE indicated the purified proteins were of suitable purity (>95%) for crystallographic analysis.

All four designed proteins were subjected to an initial crystallographic screen based on the conditions used to crystallize the wild type protein (20, 29–32). Crystals were obtained only for design BIF_1, but were needles and not suitable for x-ray crystallography. An automated high-throughput crystallization system was then used to identify new conditions (0.1 M sodium citrate, 15% PEG 6000, pH=5.5) in which large crystals of BIF_1 grew. The structure of BIF_1 was solved to 1.8 Å resolution by x-ray crystallography using molecular replacement with the parent scaffold (PDB ID 1y2q) serving as a search model. Density for BiPhe was clearly observed in a $2F_o - F_c$ map (Fig. 2A), and the torsion angle between the two phenyl rings was determined to be $\sim 28^\circ$ (Fig. 2 A). This value represented a rotation of $\sim 17^\circ$ toward the desired planar conformation relative to the dihedral found at the energetic minimum, but remained far from the desired value of 0° . In order to force the BiPhe torsion angle closer to planarity, a second round of computational design was undertaken on the basis of a detailed analysis of the structure of BIF_1.

Globally the structure of BIF_1 matched the design model quite well (Fig. 2 B, r.m.s.d. to the design model of ~ 1.3 Å over all atoms), although differences are apparent in the vicinity of the BiPhe residue. In the model, Trp42 and Trp81 both form edgewise interactions with the BiPhe side chain. However, in the crystal structure, the indole ring of Trp42 rotates such that it packs lengthwise against the BiPhe side chain in an orientation that would clash with

the side-chain orientation of Trp81 predicted by the design model (Fig. 2 C). To avoid this unfavorable steric interaction, a substantial displacement of the loop consisting of residues 81-89 likely occurs, as evidenced by the lack of electron density for residues 83-86 in the crystal structure (Fig. 2 B). In an attempt to return the disordered loop to its native position, we independently mutated each Trp to Phe (the wild type residue at both positions) in silico. Unconstrained repacking and minimization calculations in the context of each mutation showed that Trp42Phe increased SC and scored slightly better than the original design. As a result, this mutation was made standard for the remainder of the computational redesign.

A second focus of the redesign effort was to identify point mutations in residues packing against the biphenyl rings to force the side chain into the desired planar conformation. In the structure of BIF_1, the biphenyl side chain is rotated approximately 5° about chi1, and 25° about chi2 relative to its placement in the design model (Fig. 2 C). Although the phenyl ring closest to the backbone (ring A) is out of plane with respect to the model, the ring farthest from the backbone (ring B) is essentially in the plane of the design model (Fig. 2 A and C). Thus, it appears that rotation about chi2 is the predominant determinant of the 28° deviation from planarity observed in the crystal structure relative to the design model. Because Ring A is bounded on one face by the protein backbone, we believed it would be difficult to identify a mutant that would adjust this ring in the desired direction (Fig. 2 C). In contrast, ring B is flanked by both Ala79 and Tyr123, suggesting that mutagenesis of one or both of these residues could potentially planarize the BiPhe dihedral angle (Fig. 2 A). Analysis of the structure of BIF_1 suggested that the phenyl ring of Tyr123 likely prevents ring B from rotating into the plane of ring A (Fig. 2 A); thus, we mutated Tyr123 in silico to the smaller residues Ala and Val. Concurrently, Ala79 was mutated to the bulkier residues Cys, Ser, Thr, and Val. Rosetta was then used to analyze these potential sequence alternatives, again by carrying out unconstrained repacking and minimization calculations. The energies for all mutants tested fell within ~6 Rosetta Energy Units (REU) of one another, and gave SC values that differed at most by ~5%. The tight distribution of values for both metrics suggested that no clear preference exists for one mutant over another, so a series of four mutants were examined experimentally.

All four mutants expressed well and afforded diffraction-quality crystals. We solved the structures of BIF_1.1, BIF_1.2, BIF_1.3, and BIF_1.4 to 2.10, 2.50, 2.36, and 2.10 Å resolution, respectively, again by molecular replacement. In all cases the Trp42Phe mutation returned the displaced loop to its native position (Fig. S2). However, a distribution of BiPhe torsion angles between 35° and 15° was observed among the four structures (Table 2). In the majority of these cases, the change in the Chi2 angle relative to the original design remained near the value of 25° observed in the initial BIF_1 structure. This result suggested that mutations to Ala79 and Trp123 have the desired effect of rotating ring B without affecting the absolute orientation of ring A. Unfortunately, substitution of Tyr123 with the beta-branched Val and the opposing Ala79 with Ser (BIF_1.1) had the effect of rotating ring B even farther out of plane (35°) than in the original structure (Fig. 3 A). This undesired rotation was partially remedied in BIF_1.2 ($\Phi = 21^\circ$, Fig. 3 B) by substituting Ala79 with a bulkier Val residue, and further corrected in BIF_1.3 and BIF_1.4 ($\Phi = 15^\circ$ and 20° respectively, Fig. 3 C and D) by replacing the opposing Tyr123 with a smaller Ala residue and Ala 79 with Ser (BIF_1.3) or Val (BIF_1.4). This analysis suggested that ring A could

potentially be rotated into a coplanar geometry by further increasing the size of the amino acid at position 79 with an Ala79Ile mutation while maintaining Phe42 and the Tyr123Ala mutation. The additional methyl group of the isoleucine should force the side of ring A to rotate further in the desired direction.

We next generated the corresponding BIF_0 mutant (S8A, I11BiPhe, Y79I, F81W, K121I, F123A), purified the protein, and solved its crystal structure to 2.05 Å resolution (Figure 4). Analysis of the electron density showed that the two phenyl rings of BiPhe are coplanar, which matches the configuration of the TS for the bond rotation reaction. The structure of BIF_0 shows that in addition to adding steric bulk beneath ring A, the V79I mutation also forces the side chain of Phe77 to adopt a different rotamer than was observed in BIF_1.4, which has the effect of further rotating Ring A into the plane of Ring B (Fig. 4). The mutations introduced into BIF_0 do not appear to substantially affect the thermal stability of the protein. The melting temperature of this mutant, as determined by differential scanning calorimetry, was ~110°C, consistent with the 3D structure of BIF_0, which shows that the protein core is well-packed.

We have shown by iterative computational design, mutagenesis and protein structure determination that one can design a protein core that stabilizes a simple conformational transition state to such a degree that one can determine its 3D x-ray crystal structure. A similar strategy was recently employed to directly observe catalyst-substrate interactions through x-ray crystallographic analysis (33). The results described here may not be all that surprising given that enzymes typically stabilize a rate limiting TS by 8 to 12 kcal/mol. Nonetheless, these experiments underscore the ability of proteins to fold into defined 3D structures in which van der Waal's, hydrogen bonding and electrostatic interactions can be controlled with exquisite precision relative to what can be achieved with synthetic molecules.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Neil P. King and Po-Ssu Huang for helpful discussions. D.B and J.H.M. were supported by the Defense Threat Reduction Agency (HDTRA1-11-1-0041). J.H.M. was supported by NIGMS of the National Institutes of Health under award no. F32GM099210. P.G.S acknowledges support by the National Institutes of Health under award no. 2 R01 GM097206-05. The content is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health. Structures of BIF_1, BIF_1.1—BIF_1.4, and BIF_0 have been deposited in the Protein Data Bank under accession numbers 4S02 4S0I 4S0J 4S0K 4S0L 4S03.

References

1. Rose TS, Rosker MJ, Zewail AH. *J. Chem. Phys.* 1988; 88:6672–6673.
2. Polanyi JC, Zewail AH. *Acc. Chem. Res.* 1995; 28:119–132.
3. Srinivasan R, Feenstra JS, Park ST, Xu S, Zewail AH. *Science*. 2005; 307:558–563. [PubMed: 15637234]
4. Ihee H, et al. *Science*. 2001; 291:458–462. [PubMed: 11161194]
5. Pauling L. *Chemical and Engineering News*. 1946; 24:1375–1377.

6. Jencks, WP. *Catalysis in Chemistry and Enzymology*. Dover: 1987.
7. Tramontano A, Janda KD, Lerner RA. *Science*. 1986; 234:1566–1570. [PubMed: 3787261]
8. Pollack SJ, Jacobs JW, Schultz PG. *Science*. 1986; 234:1570–1573. [PubMed: 3787262]
9. Schultz PG, Lerner RA. *Science*. 1995; 269:1835–1842. [PubMed: 7569920]
10. Dwivedi S, Kruparani SP, Sankaranarayanan R. *Nat. Struct. Mol. Biol.* 2005; 12:556–557. [PubMed: 15908961]
11. Ceccacci F, Mancini G, Mencarelli P, Villani C. *Tetrahedron: Asymmetry*. 2003
12. Almenningen A, et al. *J. Mol. Struct.* 1985; 128:59–76.
13. Katon JE, Lippincott ER. *Spectrochimica Acta*. 1959; 15:627–650.
14. Carreira LA, Towns TG. *J. Mol. Struct.* 1977; 41:1–9.
15. Suzuki H. *Bull. Chem. Soc. Jpn.* 1959; 32:1340–1350.
16. Bastiansen O, Samdal S. *J. Mol. Struct.* 1985; 128:115–125.
17. Xie J, Liu W, Schultz PG. *Angew. Chem. Int. Ed.* 2007; 46:9239–9242.
18. Wang L, Brock A, Herberich B, Schultz PG. *Science*. 2001; 292:498–500. [PubMed: 11313494]
19. Eloranta JK. *Zeitschrift Naturforschung Teil A*. 1972; 27a:1652–1662.
20. Materials and methods are available as supplementary on *Science Online*
21. Razvi A, Scholtz JM. *Protein Sci.* 2006; 15:1569–1578. [PubMed: 16815912]
22. Bloom JD, Labthavikul ST, Otey CR, Arnold FH. *Proc. Natl. Acad. Sci. U.S.A.* 2006; 103:5869–5874. [PubMed: 16581913]
23. Zanghellini A, et al. New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci.* 2006; 15:2785–2794. [PubMed: 17132862]
24. McGaughey GB, Gagné M, Rappé AK. *J. Biol. Chem.* 1998; 273:15458–15463. [PubMed: 9624131]
25. Le Grand SM, Merz KM. *J. Comput. Chem.* 1993
26. Kuhlman B, Baker D. *Proc. Natl. Acad. Sci. U.S.A.* 2000; 97:10383–10388. [PubMed: 10984534]
27. Lawrence MC, Colman PM. *J. Mol. Biol.* 1993; 234:946–950. [PubMed: 8263940]
28. Chatterjee A, Lajoie MJ, Xiao H, Church GM, Schultz PG. *Chembiochem*. 2014; 15:1782–1786. [PubMed: 24867343]
29. Shimon LJ, et al. *Structure/Folding and Design*. 1997; 5:381–390. [PubMed: 9083107]
30. Dwivedi S, Kruparani SP, Sankaranarayanan R. *Acta Crystallogr. D Biol. Crystallogr.* 2004; 60:1662–1664. [PubMed: 15333948]
31. Tanaka Y, et al. *Proteins*. 2005; 61:1127–1131. [PubMed: 16224786]
32. Cosgrove MS, et al. *Biochemistry*. 2006; 45:7511–7521. [PubMed: 16768447]
33. Han S, Le BV, Hajare HS, Baxter RHG, Miller SJ. *J. Org. Chem.* 2014; 79:8550–8556. [PubMed: 25147913]
34. Hanwell MD, et al. *J. Cheminform.* 2012; 4:17. [PubMed: 22889332]
35. Fleishman SJ, et al. *PLoS ONE*. 2011; 6:e20161. [PubMed: 21731610]
36. Santarsiero BD, et al. *J. Appl. Crystallogr.* 2002; 35:278–281.
37. McPhillips TM, et al. *J. Synchrotron Radiat.* 2002; 9:401–406. [PubMed: 12409628]
38. Gonzalez A, Moorhead P, McPhillips SE. *J. Appl. Crystallogr.* 2008; 41:176–184.
39. Kabsch W. *Acta Crystallogr. D Biol. Crystallogr.* 2010; 66:125–132. [PubMed: 20124692]
40. McCoy AJ, et al. *J. Appl. Crystallogr.* 2007; 40:658–674. [PubMed: 19461840]
41. Emsley P, Cowtan K. *Acta Crystallogr. D Biol. Crystallogr.* 2004; 60:2126–2132. [PubMed: 15572765]
42. Collaborative Computational Project, Number 4. *Acta Crystallogr. D Biol. Crystallogr.* 1994; 50:760–763. [PubMed: 15299374]

Fig. 1. Strategy to capture a transition state

A. The orientations of and relative changes in energy (ΔG_{rel}) between the ground ($\Phi \approx 45^\circ$) and transition states ($\Phi = 0^\circ$ and 90°) of rotation around the central bond of biphenyl are shown. **B.** An overview of the computational design process used to generate candidate designs is shown.

Fig. 2. X-ray crystallographic analysis of BIF_1

A. The x-ray crystal structure of BIF_1 is shown in yellow; BiPhe and surrounding residues are shown in sticks. Rings A and B are those closest to, and farthest from the protein backbone respectively. Electron density around the BiPhe side chain is shown as $2F_o - F_c$ map contoured to 2σ . B. A comparison of the design model (gray) to the structure (yellow) of BIF_1 is shown; BiPhe, Trp42 and Trp81 are shown in sticks. A loop corresponding to residues 81-89 of the parent scaffold is shown in red. Missing density in the structure corresponding to residues 83-86 of BIF_1 is shown as a dashed red line. C. A comparison of the structure of BIF_1 (yellow) to the design model (gray) is shown. BiPhe, Trp42 and Trp81 are shown in sticks. An arrow indicates rotation about chi2 in the structure relative to the design.

Fig. 3. A comparison of the crystal structures of BIF_1.1-BIF_1.4

A-D. Crystal structures of second round mutants BIF_1.1—BIF_1.4 are shown. The side chains of BiPhe, and those at positions 79 and 123 are shown in sticks. Electron density from a $2F_o - F_c$ map contoured to 1.5σ (A–X) and 2.0σ ( ) is shown for the aforementioned residues. The measured dihedral angle between the two biphenyl rings is shown beneath the biphenyl side chain in each case.

Fig. 4. X-ray crystal structure of BIF_0. A

The crystal structure of BIF_0 is shown in blue and BIF_1.3 is shown in gray. The BiPhe side chain and surrounding residues are shown in sticks. B. Packing interactions between the designed protein BIF_0 and the BiPhe side are highlighted with space filling representations of the interacting residues. C-D. The structure of BIF_0 is shown, highlighting the BiPhe side chain; views from the front and side are shown. The BiPhe side chain and surrounding residues are shown in sticks and $2F_o - F_c$ maps are contoured to 2σ in each case.

Table 1

First round computational designs. Designed protein names, parent scaffolds and mutations made are listed.

Designed Protein	Parent Scaffold (PDB ID)	Computationally Designed Mutations
BIF_1	1y2q	S8A, I11BiPhe, F42W, Y79A, F81W, K121I, F123Y
BIF_2	1anu	V20G, F22I, C33W, F35BiPhe, F37W, F96G
BIF_3	2h2h	F36L, F48BiPhe, Y49W, F51W, A52M, I56G, P145G, I147A
BIF_4	1ve0	I21V, V25BiPhe, S26A, V39G, C46A, I48V

Table 2

Second and third round crystallographic analysis. Second round mutant identities, biphenylalanine dihedral angle, and x-ray crystal structure resolution are listed. Dihedrals listed are averages of those measured on each side of the biphenyl ring.

Scaffold	F42	Y79	F123	BIF Φ (deg.)	Resolution ($\text{\AA}$)
BIF_1	W	A	Y	26	1.79
BIF_1.1	F	S	V	35	2.10
BIF_1.2	F	V	V	21	2.50
BIF_1.3	F	S	A	15	2.36
BIF_1.4	F	V	A	20	2.10
BIF_0	F	I	A	0	2.05