Synthesis and folding of a mirror-image enzyme reveals ambidextrous chaperone activity

Matthew T. Weinstock1, Michael T. Jacobsen, and Michael S. Kay2

Department of Biochemistry, University of Utah School of Medicine, Salt Lake City, UT 84112-5650

Edited by Gregory A. Petsko, Weill Cornell Medical College, New York, NY, and approved July 9, 2014 (received for review June 11, 2014)

Mirror-image proteins (composed of d-amino acids) are promising therapeutic agents and drug discovery tools, but as synthesis of larger d-proteins becomes feasible, a major anticipated challenge is the folding of these proteins into their active conformations. In vivo, many large and/or complex proteins require chaperones like GroEL/ES to prevent misfolding and produce functional protein. The ability of chaperones to fold d-proteins is unknown. Here we examine the ability of GroEL/ES to fold a synthetic d-protein. We report the total chemical synthesis of a 312-residue GroEL/ES-dependent protein, DapA, in both l- and d-chiralities, the longest fully synthetic proteins yet reported. Impressively, GroEL/ES folds both l- and d-DapA. This work extends the limits of chemical protein synthesis, reveals ambidextrous GroEL/ES folding activity, and provides a valuable tool to fold d-proteins for drug development and mirror-image synthetic biology applications.

Significance

This paper addresses a fundamental question: Can natural chaperones fold mirror-image proteins? Mirror-image proteins (composed of d-amino acids) are only accessible by chemical synthesis, but are protease resistant and therefore have tremendous potential as long-lived drugs. Many large/complex proteins depend on chaperones for efficient folding. Here we describe the total chemical synthesis of a 312-residue chaperone-dependent protein (DapA) in natural (l-) and mirror-image (d-) forms, the longest fully synthetic proteins yet reported. Using these proteins we show that the natural bacterial GroEL/ES chaperone is “ambidextrous”—i.e., it can fold both natural and mirror-image proteins via nonspecific hydrophobic interactions. Our study also provides proof-of-concept for the use of natural GroEL/ES to fold d-proteins for mirror-image drug discovery and synthetic biology applications.

Author contributions: M.T.W., M.T.J., and M.S.K. designed research; M.T.W. and M.T.J. performed research; M.T.W., M.T.J., and M.S.K. analyzed data; and M.T.W., M.T.J., and M.S.K. wrote the paper.

Conflict of interest statement: M.S.K. is a Scientific Director, consultant, and equity holder of the O-Peptide Research Division of Navigen, which is commercializing o-peptide inhibitors of viral entry.

This article is a PNAS Direct Submission.

1Present address: Synthetic Genomics, La Jolla, CA 92037.

2To whom correspondence should be addressed. Email: kay@biochem.utah.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1410900111/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1410900111
Results

Selection of DapA as Model Protein. We began our investigation by searching for the smallest model protein that requires GroEL/ES for folding under physiologic conditions and has a robust activity assay that does not depend on complex chiral reagents (e.g., cofactors or other enzymes that would also require mirror-image synthesis). The E. coli DapA protein [4-hydroxy-tetradecapicolinate synthase (EC 4.3.3.7)] (Fig. 1) meets these criteria. DapA is a 31-kDa protein that forms a homotetramer (22) and catalyzes the stereoselective (23) condensation of L-aspartate-β-semialdehyde and pyruvate to (4S)-4-hydroxy-2,3,4,5-tetrahydropicolinic acid (24–26), a key step in the biosynthesis of lysine and diaminopimelic acid, a cell-wall precursor.

DapA is highly enriched in GroEL/ES complexes under normal growth conditions (8) and is aggregated or degraded in GroEL-depleted cells (8, 12). In E. coli GroEL-depletion strains, cell death occurs via lysis due to loss of DapA activity, further demonstrating the dependence of DapA on GroEL/ES to adopt its native structure (27). Indeed, in vitro, DapA is absolutely dependent on GroEL/ES for proper folding in physiologic buffer at 37 °C (8). An added benefit of DapA as a model protein is that although it depends on chaperones for folding under physiologic conditions, a chemical-folding procedure has been reported using 0.5 M arginine (8), providing an important positive control for enzyme activity independent of chaperone-mediated folding. Although a chemical refolding protocol for DapA was reported, many complex proteins cannot be folded by known chemical means, e.g., E. coli METK (28) and METF (8, 29).

Synthesis of α- and β-DapA. Because α-proteins can only be accessed through chemical synthesis, a synthetic route to DapA was devised. Synthetic peptides are routinely made using solid-phase peptide synthesis (SPPS), but a project of this magnitude (312 residues) is well outside the capability of current SPPS technology (generally ~50 residues). To access larger synthetic assemblies, chemoselective ligation techniques, especially native chemical ligation (30), are used to assemble peptide segments into larger constructs (reviewed in refs. 5, 31, and 32). Recent noteworthy synthetic proteins include tetraubiquitin [alone (33) and as part of a semisynthetic tetraubiquitinated α-synuclein (34)], covalent HIV protease dimer (35), γ-d-snow flea anti-freeze protein (36), glycosylated EPO (37, 38), and the γ-subunit of F-ATPase (39).

For synthesizing DapA, we used a recently developed method to join peptide segments via native peptide bonds formed between a peptide with a C-terminal hydrazide (for selective conversion to a thioester) and a peptide with an N-terminal Cys (40–43). We selected this chemistry because of the convenient route to peptide hydrazides via Fmoc SPPS (less hazardous and more compatible with acid-sensitive modifications than Boc chemistry), the robustness of the native chemical ligation reaction (30, 44), and the ease of carrying out convergent protein assembly (vs. linear C- to N-assembly).

Our retrosynthetic analysis began by locating all Cys residues (potential ligation junctions) in DapA (Fig. 1A), all of which are located at acceptable ligation junctions (see refs. 40 and 45 for discussions of unacceptable junctions). This information allowed us to break the protein into six segments, leaving two segments >50 residues. To expand the range of potential ligation junctions, we used a free radical-based desulfurization reaction that enables selective conversion of unprotected Cys to Ala (46, 47). This technique allows one to substitute a Cys for a native Ala residue during peptide synthesis (for use in ligation) and then convert the Cys back to the native Ala following assembly. Using

![Fig. 1](image-url)
this method, we introduced two additional junction sites at A77 and A211, resulting in eight segments overall, ranging in size from 27 to 50 residues (DapA 1–8; Fig. S1). Using optimized SPPS reaction conditions and RP-HPLC column selection (Materials and Methods), we synthesized and purified all eight peptides.

Our initial strategy for the assembly of these eight segments required 12 steps (seven ligations, two desulfurizations, and three Acm removals; Fig. S1) and their associated purifications. Acm was used as an orthogonal Cys protecting group that prevents cyclization/polymerization of peptides containing both an activated C-terminal hydrazide and an N-terminal Cys, and also prevents Cys desulfurization. Following this scheme, we assembled the C-terminal segments (DapA 5–8), but were unable to assemble the N-terminal segments (DapA 1–4). A significant complication was the His thioester on DapA 2 (H76), which was highly susceptible to hydrolysis, leading to low reaction yields during the DapA 2/3 ligation step (Fig. S2). This difficulty, coupled with the large number of manipulations (and concomitant sample losses), resulted in a failure to assemble DapA 1–4 in usable yield.

We reasoned that we could simplify the assembly if we eliminated the desulfurization step necessary to convert the Cys to native Ala at the DapA 2–3 junction. Toward this end, we determined locations in our protein that would likely tolerate permanent mutation to Cys. BLAST analysis of the E. coli DapA identified the 1,000 most-similar homologs (>69% conservation, >49% identity), which were aligned to determine positions where Cys residues naturally occur. Fortuitously, 12% of the aligned sequences contained Cys at position 77, site of the DapA 2–3 junction (Fig. 2A). Next, we analyzed the DapA crystal structure to determine the likelihood of the A77C mutation to disrupt protein structure/function. The side chain of residue 77 is surface-exposed and not in close proximity to the active site or any native Cys residues (>12 Å to the nearest Cys; Fig. 2B). This analysis suggested that introduction of the A77C mutation would likely be well tolerated. Indeed, this mutation affected neither recombinant protein activity (Fig. 2C) nor its dependence on GroEL/ES for folding under physiological conditions (Fig. 2D and Table S1). We ultimately selected a final assembly strategy that incorporated both the A77C mutation and a unified DapA 7–8 segment (we were not able to produce high-quality DapA 1–2, 3–4, or 5–6 unified peptides). This final strategy yielded a seven-segment assembly scheme (Fig. 1B) that removed four synthetic steps (and associated purifications) from the initial scheme.

Following this simplified strategy, we successfully assembled the 312-residue synthetic DapA A77C (hereafter referred to as “DapA”) in both the- and d- chiralities (Fig. 3 and SI Text), the longest synthetic peptides reported to date. The peptides were synthesized at milligram scale (1.1 and 1.7 mg of l- and d-DapA, respectively; Figs. S3 and S4). The synthetic l- and d-peptides behave identically to recombinant DapA on a C4 RP-HPLC column (Fig. 3A), and the major products possess the correct mass (Fig. 3B and C).

![Fig. 2](image_url) Validation of the DapA A77C mutation. (A) Natural sequence diversity at position 77 from Protein BLAST analysis. (B) Structure of DapA tetramer (PDB ID code 1DHP) showing, on one subunit, the surface-exposed alanine at position 77 (cyan), natural cysteine residues (green), and catalytic lysine at position 181 in the active site (red). (C) Enzyme activity of recombinant native WT and A77C DapA. Error bars indicate SD of at least three measurements. (D) GroEL/ES-mediated refolding of recombinant WT and A77C DapA.

![Fig. 3](image_url) Analysis of synthetic unfolded l- and d-DapA. (A) Analytical RP-HPLC of recombinant (black), synthetic l- (red), and synthetic d-DapA (blue) on C4 column (linear gradient 5–100% buffer B over 30 min; buffer A, 0.1% TFA in water; buffer B, 0.1% TFA in 10% water/90% acetonitrile). (B and C) LC-MS analysis of the synthetic l- and d-DapA, respectively. Observed masses were calculated using the Bayesian Protein Reconstruct tool in Analyst 1.5.1 software (AB Sciex) over the charge states covering 650–1,050 Da. See SI Appendix for larger, detailed mass spectra of the final synthetic products and HPLC and LC-MS characterization of all synthetic intermediates.
Fig. 4. Structural and functional characterization of synthetic folded L- and D-DapA. (A) Circular dichroism spectra of Arg-folded and SEC-purified recombinant (black), synthetic L- (red), and synthetic D-DapA (blue). (B) Enzyme activity of Arg-folded and SEC-purified synthetic L- and synthetic D-DapA compared with native recombinant DapA. Error bars indicate SD of at least three assays.

Our initial efforts to fold these synthetic peptides using GroEL/ES resulted in measurable enzymatic activity, albeit at relatively low levels (~20–40%; Table S1), likely due to microheterogeneity in the synthetic peptides (SI Appendix). Because active DapA assembles as a tetramer, we reasoned that we could enrich for “foldable” protein by using a chemical refolding procedure followed by size-exclusion chromatography (SEC).

Chemical-Mediated Folding of DapA. Chaperone-independent folding of DapA has been described using 0.5 M L-arginine (8), a protein refolding additive (48). This method was validated at 13 °C using recombinant DapA and works equally well with D-arginine (Fig. S5). Thus, L-arginine can be used to fold both L- and D-DapA. Importantly, this procedure also provides a chaperone-independent means to evaluate the activity of our synthetic constructs.

After arginine-assisted folding of synthetic L- and D-DapA, we isolated tetrameric protein using SEC (Fig. S6). Following SEC, both the L- and D-DapA synthetic proteins have the expected CD spectra (Fig. 4A) and are enzymatically active (Fig. 4B), demonstrating that both L- and D-synthetic proteins are correctly folded and functional. As hoped, the SEC purification generated synthetic proteins with high specific activity (~80% compared with recombinant protein). However, the Arg-assisted refolding/SEC purification resulted in a substantial (10-fold) yield loss, largely due to precipitation during dialysis and concentration steps.

Chaperone-Mediated Folding of DapA. With folded and equally active synthetic L- and D-DapA in hand, we were poised to perform the definitive experiment comparing the refolding of our synthetic L- and D-DapA by GroEL/ES. This experiment answers the question of whether GroEL/ES is ambidextrous (i.e., Can it fold a mirror-image protein?). The SEC-purified proteins were denatured for 1 h in denaturation buffer (containing 6 M GuHCl) and then diluted 100-fold into refolding buffer with or without GroEL/ES at 37 °C to initiate refolding. At specific time points, refolding was quenched by Mg chelation [1,2-diaminocyclohexanetetraacetic acid (CDTA)] followed by measurement of enzyme activity using a colorimetric assay (8). Interestingly, GroEL/ES refolded both synthetic L- and D-DapA, as demonstrated by the recovery of significant enzymatic activity (Fig. 5 and Table S1).

Discussion

The results presented here demonstrate that GroEL/ES is able to fold a D-protein and therefore does not manifest strict stereospecificity in folding its substrates. This result supports a substrate binding mechanism via nonspecific hydrophobic interactions followed by sequestration in the GroEL/ES cage (9, 10). Our study also provides proof-of-concept for the use of natural (−) GroEL/ES to fold D-proteins for mirror-image drug discovery and synthetic biology applications.

To determine if the ability of GroEL/ES to fold D-proteins is universal, the most definitive approach would be the total chemical synthesis of D-GroEL (548 residues) and D-GroES (97 residues), followed by screening of a suite of well-characterized recombinant L-substrates in refolding assays. Though we observed no difference in the activity of chemically refolded synthetic L- vs. D-DapA, there was a noticeable difference in their chaperone-mediated refolding. More detailed folding studies (49) requiring additional material will be needed to determine whether this difference reflects a general chiral preference in the recognition and/or extent/rate of folding.

Although the synthetic proteins show high specific activity (~80% of recombinant protein; Fig. 4B), it will be important to improve their quality and yield to expand application of this work to even larger synthetic proteins. We speculate that subtle synthetic defects in our proteins include single-residue deletions, racemization (50), and aspartame formation (51, 52).

Ultimately, the ability to chemically synthesize proteins of interest not only serves to advance mirror-image drug discovery efforts by making larger targets available, but also provides alluring possibilities for mirror-image synthetic biology (20) and complements efforts to synthesize other large biomolecules (e.g., synthetic genomes) (53). An intriguing prospect is the assembly of a mirror-image in vitro translation apparatus (including mirror-image ribosomal proteins in combination with mirror-image rRNAs; all but one of the 70S subunits are <300 residues), an effort that we have dubbed the “D. coli” project (18). Such a tool would not only provide a facile route to the production of mirror-image biomolecules for drug discovery, but would also facilitate the structural/biochemical study of highly toxic agents in (nontoxic) mirror-image form.

Materials and Methods

Peptide Synthesis and Ligation. Peptides were synthesized via Fmoc–SPPS on a commercial peptide synthesizer (Prelude; Protein Technologies, Inc.). Peptide hydrazides were prepared on 2-hydrazine chlorotrityl resin (ChemPep). Peptide hydrazides were activated in 6 M GuHCl, 100 mM sodium phosphate,
PhD (5-20 mM NaNO3) at −20 °C for 20 min. Peptides were then ligated in 6 M GuHCl, 200 mM 4-mercaptoethanol acid (MPAA), 200 mM sodium phosphonate, pH adjusted by HPLC of crude peptides was performed on either Phenomenex Jupiter 4-μm Protea C12 90 Å (150 × 4.6 mm) or Phenomenex Jupiter 5-μm C4 300 Å (150 × 4.6 mm) columns. Preparative reverse-phase HPLC of ligation reactions was performed on a Phenomenex Jupiter 4-μm Protea C12 90 Å (250 × 21.2 mm) or Phenomenex Jupiter 10-μm C4 300 Å (250 × 21.2 mm) column. Preparative reverse-phase HPLC of ligation products was performed on a Phenomenex Jupiter 10-μm C4 300 Å (250 × 10 mm) column. Purified peptides were analyzed by LC/MS on a Phenomenex Aeris WIDEPRO 3.6-μm C4 (50 × 2.1 mm) column on an AB Sciex API 3000 LC/MS/MS system. The major observed deconvoluted masses from mass spectrometry were calculated using Bayesian Peptide and Protein Reconstruct Tools in Analyst 1.5.1 Software (AB Sciex). See SI Appendix for full characterization of all peptides.

Enzyme Activity Assay. Ten-microliter samples of DapA (250 nM) were added to 240 μL of DapA assay buffer [6 mM imidazole (pH 7.4), 35 mM Na pyruvate, 4 mM (S)-aspartate-β-semialdehyde, 0.5 mg/mL a-aminobenzaldehyde, 12.5 mM CDTA, 1 μM DapA]. The assay was quenched after 15 min of agitation at room temperature on a microplate shaker (800 rpm) by the addition of 50 μL of 2 M HCl, developed by continuing the agitation for 1 h at room temperature, followed by measuring absorbance at 562 nm. Under these conditions, this assay demonstrates good linearity (A_{562} = 0.4 for WT recombinant DapA; saturation occurs at A_{562} > 1.5).

Arginine-Assisted Folding. DapA constructs (both recombinant and synthetic) were dissolved in denaturation buffer [6 M GuHCl, 20 mM MOPS (pH 7.4), 100 mM KCl, 10 mM MgCl2, 10 mM DTT] with 0.5 M arginine and diluted to final concentration of ~37 μM. Samples were incubated at room temperature for 40 min, 13 °C for 20 min, and then dialyzed [Slide-A-Lyzer minidialysis cassettes 3500 molecular weight cutoff (MWCO)] against 100X-volume of refolding buffer [20 mM MOPS (pH 7.4), 100 mM KCl, 10 mM MgCl2, 10 mM sodium pyruvate, 1 mM DTT] with 0.5 M arginine for 2.5 h at 13 °C. Samples were then further dialyzed against 100X:volume 100 mM ammonium bicarbonate (pH 8) for 1 h. The dialyzed sample was used directly in functional assays (post-Arg and pre-SEC) or concentrated by Vivaspin 500 10,000 MWCO centrifugal concentrators and further purified by SEC (Superdex 200 10/30; GE Healthcare) in 100 mM ammonium bicarbonate (pH 8) running buffer with a flow rate of 0.75 mL/min (post-Arg and post-SEC). Following SEC, samples were again concentrated and prepared for structural (CD spectroscopy) and functional assays (direct activity and GroEL/E8 refolding).

Chaperone Refolding Assay. The DapA refolding assay (to evaluate GroEL/E8 chaperone activity) was adapted from ref. 8. Twenty-five-micromolar stocks of DapA were prepared from lyophilized powder (pre-SEC) or buffer exchanged (post-SEC) into denaturation buffer [20 mM MOPS (pH 7.4), 100 mM KCl, 10 mM MgCl2, 10 mM sodium pyruvate, 5 mM ATP] with or without 7 μM GroEL monomer and 7 μM GroE monomer. Final DapA concentrations used in refolding assays were 250 nM. At specific time points, 10-μL aliquots of the refolding reaction were added to 240 μL of DapA assay buffer, which simultaneously quenches chaperone-mediated refolding and initiates assay of enzyme activity (as described above).

ACKNOWLEDGMENTS. We thank Costa Georgopoulos and Debbie Ahe for their helpful discussions, provision of protocols, gift of our initial stocks of GroEL, and critical reading of the manuscript; Rob Marquardt for his mass spectrometry advice and training; the Gary Keck laboratory for use of its ozone generator for preparing (S)-aspartate-β-semialdehyde; Janet Iwasa for advice on figure design; and Debra Eckert for critical review of the manuscript. Funding for this work was provided by National Institutes of Health (NIH) Predoctoral Training Grant F31CA171677 (to M.T.J.) and NIH Grants AI076168 and AI102347 (to M.S.K.).


