Size Matters: Side Chain Length Affects SH2 Substrate Binding

Christian F.W. Becker[1,*]
[1]Center of Integrated Protein Science Munich and Department of Chemistry, Technische Universität München, Lichtenbergstr. 4, 85747 Garching b. München, Germany
*Correspondence: christian.becker@ch.tum.de
DOI 10.1016/j.chembiol.2010.03.001

The incorporation of lysine analogs in Src Homology 2 (SH2) domain variants by Virdee et al. (2010) reveals the length-dependent contribution of a single amino acid side chain in recognition of phosphotyrosine-containing substrate peptides and a change in specificity with shorter side chains.

Access to large quantities of homogeneous protein preparations by recombinant expression is a major driving force behind quickly accumulating biological knowledge. However, certain biological problems require access to proteins with modified amino acids, which can be made available by modern protein engineering techniques. Development of these techniques has made dramatic progress over the last sixteen years, due to new approaches that give access to semisynthetic proteins consisting of chemically prepared peptide segments and of biologically generated protein segments. Native Chemical Ligation (NCL) (Dawson et al., 1994) has become the key reaction, not only for total chemical synthesis of proteins of up to 200 amino acids, but also for access to large semisynthetic proteins. It converges nicely with intein technology that generates access to large recombinantly expressed protein thioesters. The combination of synthetic peptides and recombinantly expressed protein segments is termed Expressed Protein Ligation (EPL) (Muir et al., 1998) and has been stretching the size limit of the proteins that are accessible to selective chemical modifications far beyond anything possible before introducing these new tools into chemical biology.

Protein semisynthesis is complemented by methods that reprogram the genetic code to incorporate unnatural amino acids either by misacylating specific tRNAs, thereby leading to a residue-specific exchange, or by suppressing a stop codon with an orthogonal tRNA/tRNA synthetase pair that allows site-selective amino acid exchange (Wiltschi and Budisa, 2007; Wang and Schultz, 2004). Such extensions of the genetic code are highly valuable but they do not offer complete freedom in terms of functional groups to be incorporated, as well as in terms of size. These limitations are mainly due to inherent properties of the almost perfect biosynthesis machinery used here, and require engineering of this complex machinery itself to be overcome.

Multiple examples exist in which scientists have applied these reactions to introduce posttranslational modifications such as lipidation, phosphorylation, acetylation, methylation, glycosylation, and even ubiquitylation into proteins (Hackenberger and Schwarzer, 2008). Such modifications or combinations thereof are otherwise difficult to generate, and homogeneous preparations are needed to elucidate their impact on protein function. These eye-catching applications of protein semisynthesis often obscure the view on similarly important opportunities amenable through these techniques: the chance to understand protein function by introducing only very subtle changes into proteins of interest. Such small changes help to illustrate the contributions of single residues, functional groups, or even atoms in binding events or catalysis. Virdee et al. (2010) provide a striking example of the effect of the length of a single amino acid side chain on substrate binding in an SH2 domain. They take advantage of the chance to introduce any unnatural amino acid into a short peptide of only 12 residues during solid-phase peptide synthesis and produce the remaining parts of the SH2 domain by recombinant expression in *E. coli*. Their approach elegantly combines the freedom and ease of chemical synthesis of a small peptide segment comprising only 12 amino acids in the middle of the SH2 domain with recombinant expression of the larger N- and C-terminal segments as shown in the retrosynthetic analysis of the synthesis strategy in Figure 1. Subsequent NCL reactions give functional SH2 domains that contain mutations of a specific lysine residue, previously described as important for substrate binding, to three unnatural amino acid buildings blocks with progressively shorter side chains but maintaining the terminal amino group.

Even though semisynthesis of the SH2 domain appears straightforward, several issues related to the chosen EPL strategy had to be solved. Some of these issues were based on the chemical demands of NCL reactions. NCL requires a sulfhydryl group for an initial thioester exchange reaction and ideally the formation of a five- or six-membered transition state on the way to a new peptide bond between two protein segments. In only a few cases, cysteine, as the second rarest natural amino acid (Trinquier and Sanejouand, 1998), is located in strategically interesting places. Therefore semisynthetic strategies based on EPL often require cysteine mutations or more elaborate approaches to circumvent such mutations. Virdee et al. (2010) had to introduce a total of four mutations into the SH2 domain, excluding the central residue of their study, lysine βD3, to enable efficient semisynthesis via a three segment strategy. Three of these mutations were necessary to satisfy the chemical needs of NCL. Such mutations are often required when designing EPL strategies and entail the need for subsequent experiments demonstrating that these mutations do not interfere with any of the properties that will be investigated.

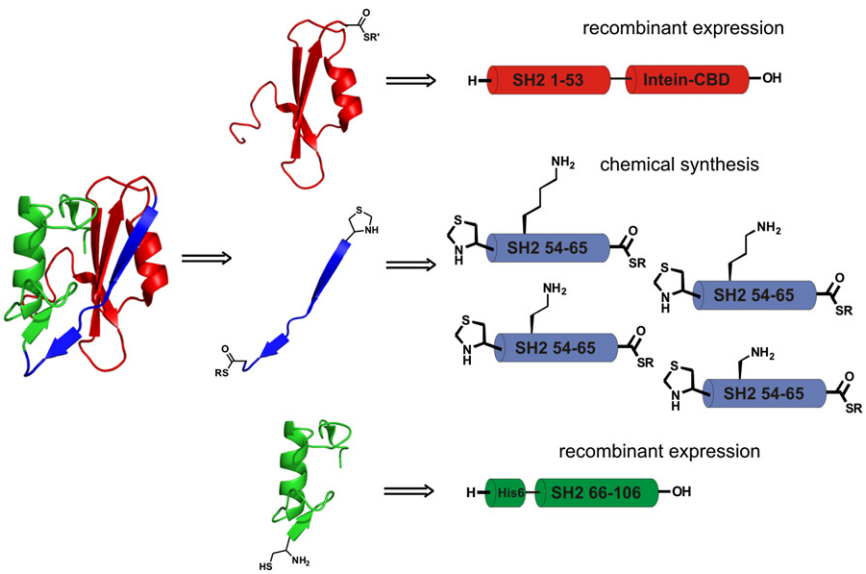

Figure 1. Retrosynthesis of Modified SH2 Domain Variants (PDB entry 3EAC)
The 106 amino acid comprising SH2 domain was divided into three segments. Four variants of the small middle segment SH2 54-65 (blue) were prepared by peptide synthesis containing lysine and three unnatural lysine analogs (L-ornithin, L-diaminobutyric acid, L-diaminopropionic acid). Longer segments were accessed via recombinant expression in *E. coli*. For SH2 1-53 (red), an intein fusion tag equipped with a chitin-binding domain (CBD) was employed for generating the required C-terminal thioester and for purification, respectively. The C-terminal segment SH2 66-106 (green) was also generated by expression in *E. coli* and purified by taking advantage of an N-terminal hexahistidine tag. Treatment with CNBr led to C-terminal cleavage after a methionine residue and release of SH2 66-106 with an N-terminal cysteine residue for NCL.

In case of the semisynthetic SH2 domain, three mutations had no effect on substrate specificity and affinity toward a reference-binding peptide (pYEEI). However, a fourth mutation, required to abolish a side reaction lowering ligation efficiency, did increase the affinity of this binder by a factor of eight. To obtain sufficient amounts of semisynthetic SH2-domain variants, this mutation was inevitable and emphasizes problems related to doing chemistry on complex macromolecules when even conservative mutations can lead to large changes in the biological function of the protein.

The SH2-domain variants in which a single lysine residue in position βD3 was replaced by three unnatural lysine analogs that differ in length by one methylene group provide a tool to probe the influence of side chain length on a rather complex network of electrostatic interactions. The initial hypothesis of Virdee et al. (2010) that binding of the model substrate pYEEI would only be effected via one key electrostatic interaction between lysine βD3 and glutamic acid +1 in pYEEI turned out to be far too simple. The complexity of these interactions is only partially revealed, since internal electrostatic effects take part in the difference in binding behavior of SH2 toward pYEEI. Previous studies with lysine analogs have revealed the influence of side chain length on electrostatic interactions between helixes (Cheng et al., 2007). However, conclusive evidence showing why nature uses the longest side chain in its native amino acid repertoire to present a primary amino group is still not available. One might speculate about changes in pKa value that coincide with side chain length, as well as accessibility for posttranslational modifications by a multitude of different enzymes (Walsh, 2006).

Reducing the side chain length by two methylene groups leads to the most surprising finding of this work: an increase in binding affinity to pYDEI but not to pYEEI. The interplay of reducing side chain length and the complexity of electrostatic interactions is not fully understood. A complete picture will only be accessible when using a combination of structure analysis and modeling. Structure analysis of semisynthetic proteins is often hampered by the amounts that can be generated. In case of the SH2 domain, sufficient amounts for structural analysis either by X-ray crystallography or NMR should become available. The opportunities, especially for NMR, to selectively label the semisynthetic SH2 domain are manifold and should allow targeting of the specific interactions with site- or even atom-selective labeling. The minimal peptide synthesis requirements for the semisynthetic strategy applied by Virdee et al. (2010) will increase the demand by biologists and could lead to SH2 domains with non-native residues that exhibit new properties useful in biochemical and biomedical applications.

REFERENCES

Cheng, R.P., Girinath, P., and Ahmad, R. (2007). Biochemistry *46*, 10528–10537.

Dawson, P.E., Muir, T.W., Clark-Lewis, I., and Kent, S.B.H. (1994). Science *266*, 776–779.

Hackenberger, C.P., and Schwarzer, D. (2008). Angew. Chem. Int. Ed. Engl. *47*, 10030–10074.

Muir, T.W., Sondhi, D., and Cole, P.A. (1998). Proc. Natl. Acad. Sci. USA *95*, 6705–6710.

Trinquier, G., and Sanejouand, Y.H. (1998). Protein Eng. *11*, 153–169.

Virdee, S., Macmillan, D., and Waksman, G. (2010). Chem. Biol. *17*, this issue, 274–284.

Walsh, C.T. (2006). Posttranslational Modifications of Proteins (Greenwood Village, CO: Roberts & Company).

Wang, L., and Schultz, P.G. (2004). Angew. Chem. Int. Ed. Engl. *44*, 34–66.

Wiltschi, B., and Budisa, N. (2007). Appl. Microbiol. Biotechnol. *74*, 739–753.