T-cell receptor signal sequences

Future studies of the physical and structural chemistry of the T-cell receptor (TCR) will be greatly facilitated by prokaryotic, or other heterologous, expression systems. In such studies, no matter whether only the mature portion is expressed and refolded or a prokaryotic signal sequence is fused, the assignment of the original signal cleavage site becomes crucial. The exact length of the TCR is also important in all model-building studies (e.g. see Ref. 6).

However, the position of signal cleavage sites of TCRs is not yet known. Therefore, most publications (e.g. Refs 3,7–12) use the alignment to antibody variable domains as a guide, or relate it to other TCRs, themselves assigned by alignments. In contrast to the older literature, which acknowledges this fact, current sequence databases do not explicitly indicate the hypothetical nature of such approaches. Thus, the signal cleavage sites have wrongly assumed the status of fact. Furthermore, the assignment of the signal cleavage site is not always reasonable from a biochemical point of view.

Using biochemical and statistical criteria13, some of the cleavage sites proposed in the literature, which would give rise to very long signal sequences, appear very implausible. Briefly, a typical signal peptide as determined experimentally for many proteins comprises three parts14: a short (1–3 residues) positively charged N-terminal region (which, in eukaryotes, includes the α-amino group as a positive charge); followed by a hydrophobic core (7–15 residues); and completed by the C-terminal signal peptide cleavage site (3–7 residues). Predictive schemes for cleavage sites have been developed using data from proteins with experimentally determined cleavage sites14. Using the algorithm of the program SIGSEQ2, cleavage sites were predicted for the α chains of six TCRs with long signal sequences in the database (Fig. 1). The sites predicted by SIGSEQ2 were found to be different from the ones given in the literature based on antibody homology, which receive exceptionally poor scores in the prediction. Even a superficial application of the biochemical criteria makes these sites improbable. In every case, cleavage would occur too far from the hydrophobic core.
of the signal peptide (Fig. 1). For this reason, our studies on the expression of TCR fragments in *Escherichia coli* utilize signal cleavage sites based on such biochemical and statistical criteria.

Independently of the considerations raised above concerning expression systems for soluble, recombinant TCRs (reviewed in Ref. 15), reconstructing the correct N-terminus of the protein will also be important. By pointing out the apparent contradiction between the widely used signal sequence assignments, and the biochemical and statistical plausibility of such alignments, we hope to provoke experimental determinations of dubious cleavage sites.

**Christoph Wülfing**  
*Universität Zürich,  
Biochemisches Institut, Winterthurerstr. 190,  
CH-8057 Zürich, Switzerland.*

**Andreas Plückthun**

Sequence similarity between β-cell autoantigens

Most humans and nonobese diabetic (NOD) mice with insulin-dependent diabetes mellitus (IDDM) manifest autoimmune reactivity to insulin and/or isofoms of the enzyme glutamic acid decarboxylase (GAD-65 and -67). Whether any of these molecules actually trigger or drive a pathogenic immune response against pancreatic β cells remains controversial, and no basis for association among them has previously been proposed. We have identified a 13 amino acid sequence of high similarity (with six identical and three conserved residues) between human GAD-65 (residues 506–518) and human pro-insulin (24–36). This similarity extends to human GAD-67 and both of the mouse pro-insulins and GADs.

Other than the similarity between pro-insulin and GAD, this sequence has several interesting features. It lies within a region of human GAD-65 (473–543) shown to be immunodominant for T cells of IDDM patients, and largely overlaps one of the two mouse GAD-65 peptides (509–528) reported to be the first recognized by NOD mouse T cells. The sequence contains a (+IXXXI−) motif characteristic of most I-Aε (I-Aνο)-binding peptides reported to date, as well as an XYYYYYYYYYYY motif for H-2Kb (Ref. 7), the major histocompatibility complex (MHC) class I of NOD implicated in the pathogenesis of IDDM (Ref. 8). Both of these motifs overlap the cleavage site in pro-insulin (Arg32–Glu33), which separates the C-peptide from the C-terminus of the insulin B chain, and would therefore be absent from the mature insulin molecule.

Evidence exists for T-cell epitope mimicry and "determinant spreading" in other autoimmune disorders. Reactivity to this pro-insulin

---

**References**