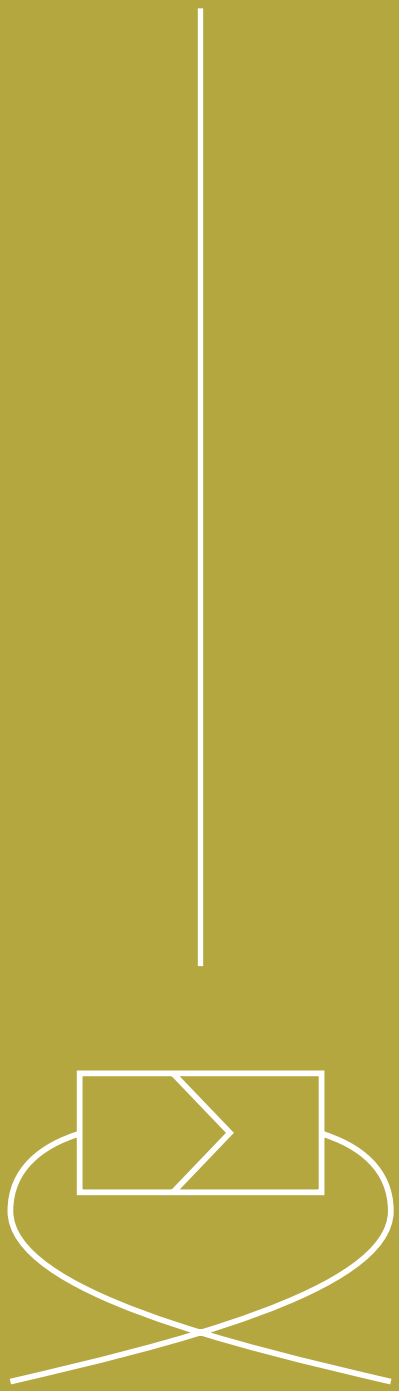
# SUPPORTING MATERIAL



STEPHEN KENT: INVENTING SYNTHETIC METHODS TO DISCOVER HOW ENZYMES WORK
 HTTPS://DOI.ORG/10.47261/1549

PROCEEDINGS FROM THE
PHARMACEUTICAL
MANUFACTURERS
ASSOCIATION'S

# SYMPOSIUM ON BIOTECHNOLOGY AND HUMAN HEALTH CARE

OCTOBER 11–14, 1987

* SUPPORTING MATERIAL

**Aug. 18th, 2022**
In October 1987, I gave a presentation to the American Pharmaceutical Manufacturers Association (AmPhar) national symposium on biotechnology and health care. It summarized the work I had done in the previous four and half years at Caltech. In many respects, that talk encompassed techniques that formed the foundations of the field we now call “Proteomics”. *Stephen Kent.*

# Protein Chemistry and the Development of Diagnostics and Therapeutics

STEVEN KENT, PH.D.

*Division of Biology, California Institute of Technology*

I thought I would talk abit about protein chemistry and, in particular, given the nature of the audience, protein chemistry and the development through biotechnology of new therapeutics, diagnostics, and vaccines. This talk is broken down into two parts. The first part will be a discussion of the application of synthetic peptides and protein chemistry to hepatitis B and interleukin-3 systems. Then, in the second part, I will discuss the use of protein sequencing and an approach to the systematic development of human diagnostics.

First of all, the underlying technology. Unlike everyone else here, I'm not a cloning person, though I do have postdocs working with me who carry out the standard techniques of molecular biology. The main underlying technology behind my research is the use of the chemical synthesis of peptides and proteins. On the surface, this appears to use fairly standard methods, as originally developed by Bruce Merrifield some twenty five years ago. My own contribution has been in effect, the development of new insights into the chemistry of solid phase synthesis that have allowed the first effective automation of the method. We developed a new instrument several years ago that is called the Automated Peptide/Protein Synthesizer. The unique thing about it is not only were we able to carry out very rapid chemical synthesis of peptides, but also the chemical synthesis of long polypeptide chains that constitute proteins. Although the polypeptide chain is common to both peptides and proteins, the difference being that in proteins the peptide chain is folded to a stable three dimensional conformation. The chemistry of assembly is somewhat different, because the demands for the synthesis of a 100 or 150 amino acid long peptide chain are much more stringent than for the chemical assembly of a 30 residue peptide chain.

However, the Automated Peptide/Protein Synthesizer can accommodate both sets of chemistry, simply by changing a floppy disk, which contains the operating system. (see Figure 1). Just to show off a bit, we've improved the chemical synthesis of peptides recently (this is a paper we submitted on the second of October.) We've got 20 minutes per residue now for the assembly of peptide chains up to about 50 residues (see Figure 2). The chemistry for the assembly of the long peptide chains found in proteins is shown later in the talk.

Let's look now at hepatitis B. I'm sure you all know that hepatitis B is a major, world wide health problem. There are several hundreds of millions of people infected worldwide, primarily in the African Asian populations of the Third World. Among the most susceptible populations are infants both in Asia and Africa, but there are differences in the epidemiology of transmission, from infected mother to infant, in these two continents. However, the problem is the same in the sense that the several hundred million infected people act as a reservoir for the disease which is then repeatedly brought into the U.S. through the standard transmission routes that apply to this and other virus diseases.

*Figure 1*

**AUTOMATED PEPTIDE PROTEIN SYNTHESIZER:**

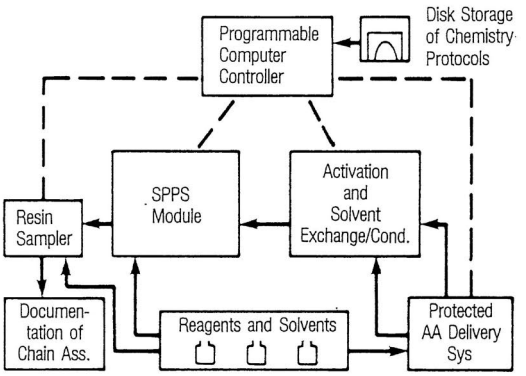



*Figure 2*

**AUTOMATED SYNTHETIC CYCLE FOR ADDITION OF A SINGLE AMINO ACID**

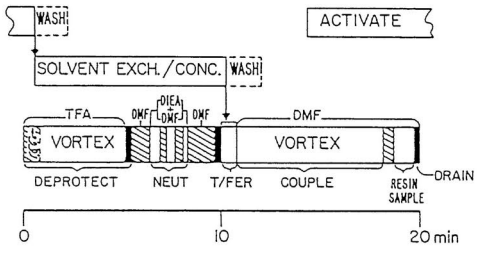


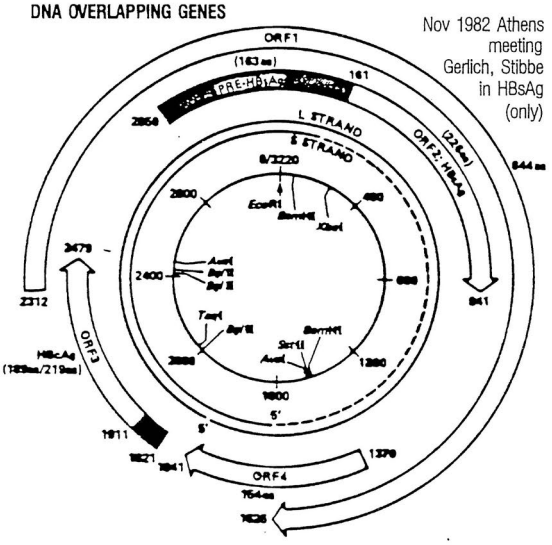

Figure 3 depicts the genetic structure of the hepatitis B virus. The sequence of the hepatitis B viral genome has been known since 1979, and several different strains have been sequenced by different research groups. It is a very compact genome, only 3.2 kilobases of DNA. We will focus on a single gene to illustrate the application of protein chemistry to the development of new products, the open reading frame shown in Figure 4, the envelope gene open reading frame, commonly referred to as the S gene, but we now know, as we will see in a moment, that there are three gene products that are produced that cover the entire open reading frame.

The hepatitis B virion itself has a typical nucleocapsid core structure and is an enveloped virus containing three different protein species in the viral envelope. As I've mentioned, the envelope gene, which codes for an open reading frame of 400 amino acids, gives rise to three different protein products, schematically represented in Figure 5.

*Figure 3*

**HBV IS ENVELOPED VIRUS 3KB GENOME-DOUBLE STRANDED DNA OVERLAPPING GENES**

Nov 1982 Athens meeting Gerlich, Stibbe in HBsAg (only)




2. Major envelope protein translation product of "S" gene: most vaccine work focused on this protein.
3. Nov. 1982 Stibbe and Gerlich presented evidence suggesting HBsAG 22nm particles contain minor protein components coded for by Pre-S.

*Figure 4*

**HEPATITIS B VIRUS "env" GENE**

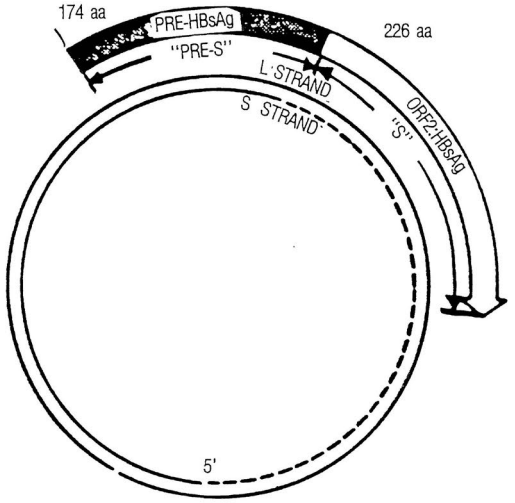
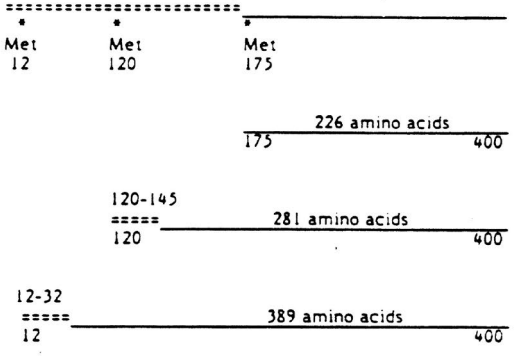


*Figure 5*

**OPEN READING FRAME OF THE ENVELOPE GENE OF HBV (TOP) AND THE THREE POSSIBLE TRANSLATION PRODUCTS: "S-PROTEIN", 175–400; "MIDDLE"-PROTEIN, 120–400; AND "LARGE"-PROTEIN, 12–400 (OR 1–400).**

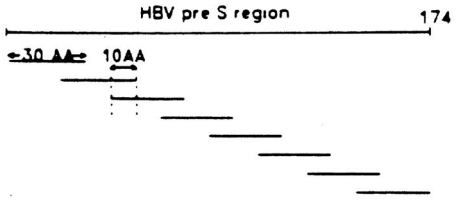



What I am going to do now is go through the use of synthetic peptide chemistry to map the epitopes coded for in the pre-S region of the envelope proteins. What I'm talking about is not the small protein but the parts that are unique to the middle protein and the large protein. This is a region of 55 amino acids, and the other region in the figure is 120 to 108 amino acids depending on the subtype of the virus.

One way that we've taken to locate interesting structures on these viral proteins is to synthesize a family of overlapping peptides. If one were totally organized and had unlimited resources, this would be done in the following way (Figure 6): for this 174 amino acid pre-S region you would make 30 amino acid peptides overlapping by about 10 amino acids, because

*Figure 6*

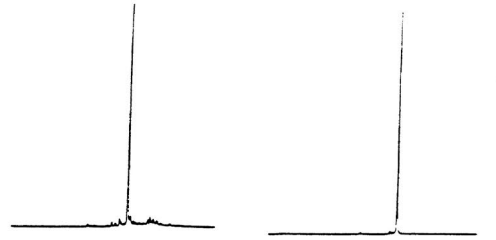
**SYSTEMATIC APPROACH TO EPITOPE LOCATION**
***Use of Large (approximately 30 AA) Peptides**




FEATURES
- Reduces Immunoassay Screening
- Stable 2° Structure
- T- and B-Cell Epitopes in Single Peptide

we know, or at least think we know in this field, most epitopes in the immune system are 10 amino acids or less. Then you would assay these peptides for their ability to bind antibodies in immune sera, for example, or compete with virus binding to receptor sites. In fact, we have more or less done that for the pre-S region of hepatitis B. Figure 7 shows one of the peptides, a 27 residue peptide fragment that actually corresponds to the amino terminal of the middle protein of the hepatitis B viral envelope.

*Figure 7*



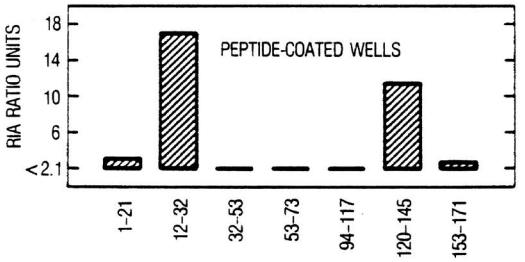
Synthetic peptide, of amino acid sequence MQWNSTAFHQTL-QDPRVRGLYLPAGG-Cys-$NH_2$, corresponding to HBV env-gene (120–145). Reverse phase HPLC (detection at 214nm) showing total crude products (left) and purified peptide (right). The two minor peaks in the purified product are from cleavage at $Asp^{133}$–$Pro^{134}$ on standing in dilute acetic acid.

We made a family of such peptides, not covering the entire pre-S region of the envelope protein or proteins quite as systematically as I showed you in the diagram. But we did make a family of peptides, covering this protein sequence. We then took pooled human immune sera from 26 individuals who had recovered from hepatitis B infection, and we asked whether there were antibodies in the pooled sera that would recognize any of these synthetic peptides (Figure 8). The answer was, of course, there were such antibodies. Two peptides were recognized very well, although there was some recognition of other peptides. Now these turn out to be the amino terminal sequences of the middle and large proteins of the hepatitis B envelope (see Figure 5).

What we are going to do now is focus in on the 26 amino acid long peptide from the amino terminal of the HBU envelope middle protein (Figures 5,7,8). The question is: is there a structure within this peptide that accounts for all the binding of the inhuman immune sera antibodies? And, again, the answer is yes (or I wouldn't be showing the example). We used several different methods, all of which gave the same answer. Each method used synthetic peptides of one sort or another. One approach, which I am not going to describe in detail, consists of making the full length peptide and substituting alanines at all the positions across the sequence and seeing what

*Figure 8*

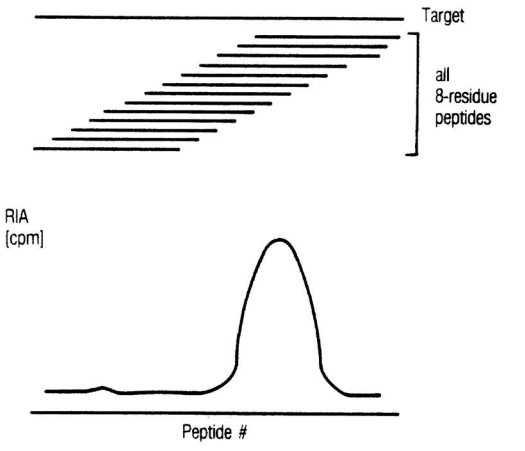
**CROSS-REACTIVITY OF LARGE PEPTIDES WITH POOLED HUMAN IMMUNE SERA**




SYNTHETIC PEPTIDES SPANNING INDICATED RESIDUES OF THE pre S PORTION OF THE HBV env PROTEIN

effect this has on the binding. The approach I am going to describe (see Figure 9) consists of making a short family of short peptides, each eight residues in length, that are offset by one residue at a time, and walk across the sequence of the larger peptide that we know binds antibodies in human immune sera.

*Figure 9*




Rather than use the synthetic technology for making free peptides that I have already told you about, we developed a way of making peptides chemically on glass fiber filter paper in the wells of microtiter plates, and that is schematically illustrated in Figure 10. We just chemically derivatized the glass fiber filter paper disks, and grew the peptide by fairly standard sold phase chemistry, making several copies of each one. After the chemistry, you don't cleave the peptide off the glass fiber disk. As shown in Figure 11, it is pretty obvious that most of the antibody binding is accounted for by a region between residues 129 and 139 in this sequence (see Figure 5) (remembering that the residues covered in this experiment were from

*Figure 10*

**PEPTIDE SYNTHESIS ON GLASS FIBER DISCS**

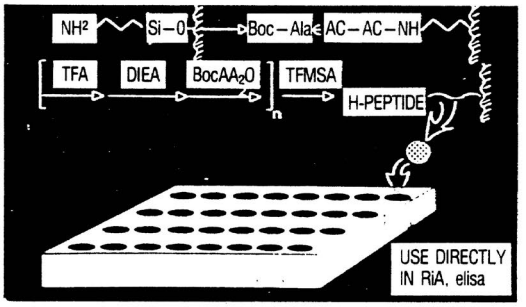
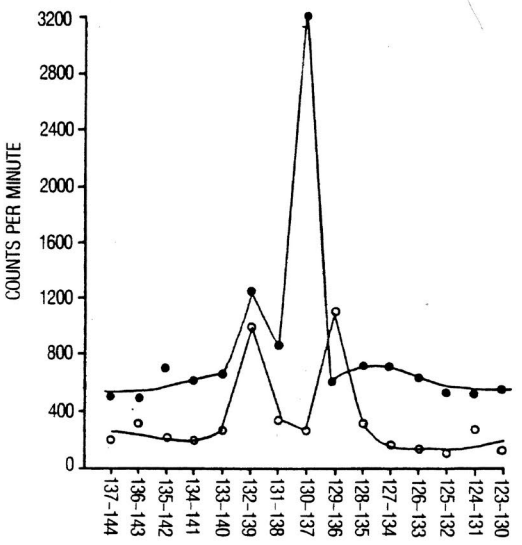



*Figure 11*

**CROSS REACTIVITY OF HUMAN IMMUNE SERA WITH 8-RESIDUE PEPTIDES SPANNING THE SEQUENCE 120-145**

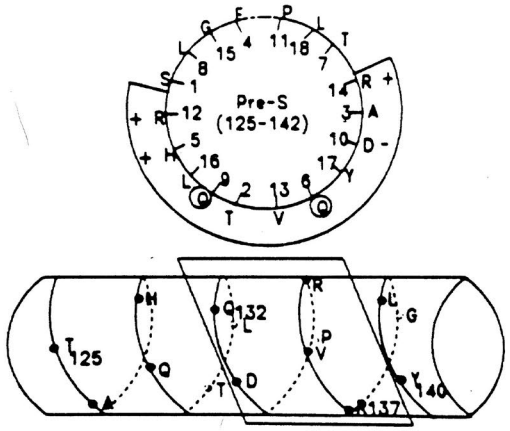



*B-CELL EPITOPE @ RESIDUES 132-137.

120 to 153). So the antibody binding is in the middle of the sequence. Showing this schematically in Figure 12 what we have is an amphipathic helix. If we show the middle part of the long peptide as an Edmondson wheel projection down the alpha helix, we see that one face is polar and one face is apolar. And if we look back at Figure 11, there were several positive points, small peptides that bound most of the antibodies. From that you can deduce that the common structure in

*Figure 12*

**B-CELL EPITOPE (RESIDUES 132-137)**
**FORM CHARGED PATCH ON SURFACE OF AMPHIPATHIC HELIX**



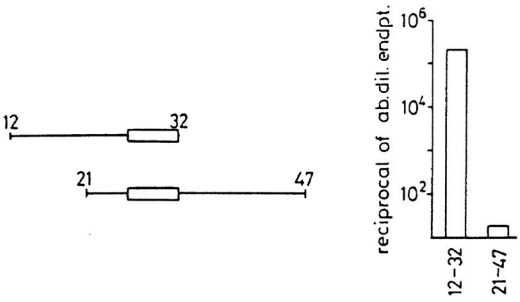

those sets of peptides was this patch of amino acids, which consists of 5 amino acids, three of which are charged, an aspartic acid and two arginines. All these residues appear on one face of this amphipathic helix (Figure 13). This has become in the last few years a fairly standard sort of thing to do with viral coat proteins or viral envelope proteins. This type of information, from this or other approaches has produced knowledge about a number of systems both in our own lab and in others.

In principle you can go ahead and design much smaller antibody binding B-cell epitopes. One very interesting thing that we have found, in the context of developing vaccines, is as follows. If we go back to our parent peptide, residues 120-145 of the hepatitis B virus envelope gene open reading frame the N-terminal 26 residues of the envelope middle protein, we have located the antibody binding site between residues 132-139. Now, if we take two peptides that each contain that antibody binding site we find a very interesting difference in their behavior. The one on the top in Figure 13 is an extremely powerful immunogen, even when not conjugated to any carrier protein; these are free peptides injected into animals without KLH or BSA or albumin - whatever is your favorite T-cell help-inducing carrier pr . On the other hand (Figure 14), this other peptide residues 128-153 while it still displays the same antibody binding activity, is virtually nonimmunogenic compared to the peptide 120-145.

*Figure 13*

**IMMUNOGENICITIES OF IDENTICAL LENGTH PEPTIDES CONTAINING THE SAME B-CELL EPITOPE**

**L-specific T cell epitope**

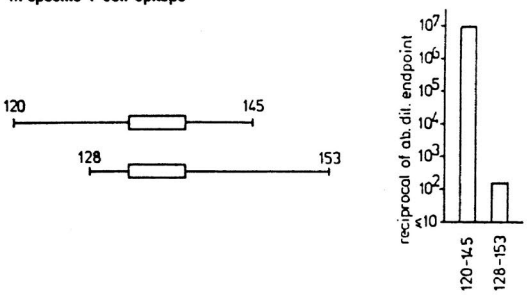



The same sort of analysis applies to the hepatitis B virus large protein. As you recall from Figure 8, the other peptide that bound antibodies in human immune sera was the amino terminal 21 residue peptide, from the large protein as shown in Figure 13. Again we find that this peptide, residues 12-32, is an extremely potent immunogen. This other peptide, offset within the large protein, still contains the antibody binding site and is fully active as an antigen, but is a very poor

*Figure 14*

**IMMUNOGENICITIES OF IDENTICAL LENGTH PEPTIDES CONTAINING THE SAME B-CELL EPITOPE**

**M-specific T cell epitope**



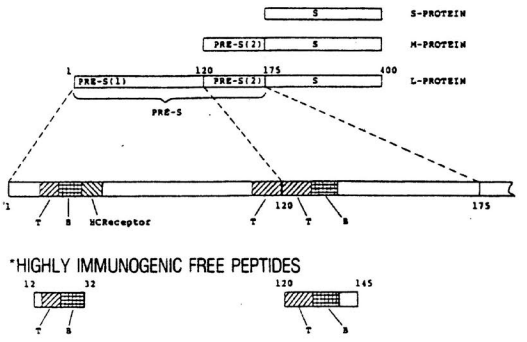
immunogen. And of course what that means is that these peptides, when they are in the right region, not only contain the antibody binding site, but also a structure that will enable the antigen presenting cell to induce T-cell help, a so-called "T-cell epitope".

So, we have a pair of synthetic peptides from two of the proteins in the envelope of the hepatitis B virus that contain not only immunodominant antibody binding sites, but also contain T helper cell determinants for those same sites. That is an interesting observation that can be used as the basis for the rational design of vaccines, at least in principle.

Taking the same sort of approach, we've mapped the epitopes in the pre-S region of the hepatitis B virus. I've already discussed these two B and two T cell epitopes; there is at least a third T cell epitope that is present, and I won't go into this linear structure that is actually the dominant component of the hepatocyte membrane receptor binding structure on the virus. It turns out though, that a linear peptide synthesized corresponding to this region of the sequence of the large protein, will actually compete with hepatitis B virus for binding on the hepatocyte. The results obtained using this synthetic peptide approach to map the epitopes of the surface-exposed regions of the HBU envelop proteins are shown in Figure 15.

*Figure 15*

**EPITOPES IN PRE-S-CODED DOMAINS OF HBV ENVELOPE PROTEINS**




That's a little bit about how we use synthetic peptide chemistry in a biologic system of this type, to ask and answer certain interesting questions. We now know, as a result of asking those questions, that the pre-S coded domains of hepatitis B virus (which incidentally were known to exist three years ago) actually have several important biological functions, as well as being targets of the cellular, and humoral immune response. The therapeutic implications of that (which I've summarized in Table I) are somewhat controversial, to put it mildly.

*Table I*

**THERAPEUTIC IMPLICATIONS**

- **EXISTING VACCINES CONTAIN ONLY S-PROTEIN**
- **STRONG, EARLY HUMAN IMMUNE RESPONSE TO PRE-S PROTEIN**
- **ANTI-(PRE-S) RESPONSE IS VIRUS-NEUTRALIZING AND PROTECTIVE**

The existing first and second generations hepatitis vaccines do *not* contain the pre-S coded regions of the envelope protein; they contain only the S protein. They are effective vaccines, as far as we know, although there's controversy about that also. But certainly in the U.S. non-immuno suppressed population, they are effective vaccines. However, we now know that in animal models, the presence of these pre-S coded sequences convert non- responder stains to responders to the S protein. There is abundant evidence now that we ought to have a look at whether we can make vaccines containing these pre-S coded domains.

The original hepatitis B vaccine consists of 22 nanometer sub-viral particles, isolated from the sera of so-called "non-infective" carriers. The second generation vaccine is the cloned equivalent of that. These do *not* contain the middle and large envelope protein domains. So, instead of having this colorful repertory of proteins in their envelope, they have only this dull repertory—namely, this single S protein. So, what can we conclude from that? Although I haven't gone into it today, (see Table II) we can conclude, as our own lab and others have shown, that there is a strong early immune response to the pre-S

*Table II*

**CONCLUSIONS**

- **Diagnostics based on these immunodominant epitopes of the novel envelope proteins may be clinically useful.**
- **Existing hepatitis vaccines should contain the full range of eptitopes present in the virion.**
- **"Second generation" vaccines should contain all biologically significant HBV envelope protein epitopes.**

Stephen Kent and Robert Neurath
Division of Biology Lindsley F. Kimball
Caltech Research Institute
New York Blood Center

coded sequences that can be used as a basis for alternative hepatitis B diagnostics. It is not clear whether they will be superior to the existing ones since the clinical correlations have yet to be worked out. We can make a case, as I've already indicated, that we should consider an improved generation of hepatitis B vaccines containing these other important epitopes.

How would one go about making these? The answer is, I hope, chemically. On the other hand, there's no reason why you can't make a synthetic vaccine by cloning for production purposes. We and others have published routes for making expression products containing multiple copies of a small synthetic peptide in such a way that when cleaved either enzymatically or chemically, it gives rise to the epitope you would like to have in a synthetic vaccine. And, in the case that we described, it would actually give rise to a chemically activated peptide ready for hooking to a carrier to make a vaccine. The chemistry route is not the only way.

I though I would summarize (as shown in Table III) what we're up to in the technology of vaccine design and production. The conventional vaccines, of course, are attenuated or inactivated pathogens. Sub-unit vaccines, which we've heard about this morning, originally were purified from natural sources and are now cloned. We are now entering the era of designed vaccines where we try and understand the basis of the immune response to a particular pathogen, and then design a structure or a molecule that will stimulate this response. And these can be produced by cloning or chemical synthesis. But this is only a way station on the route to the ultimate goal, which is a pharmaceutical type vaccine, one that would be much easier to administer to large populations, etc.

*Table III*

**EVOLUTION OF VACCINE DEVELOPMENT**

1. **Conventional Vaccines**
   - **Attenuated* Or Killed* Pathogens**
2. **Subunit Vaccines**
   - **Purified* From Natural Sources**
   - **Clones***
3. **"Designed" Vaccines**
   - **Protein Or Peptide**
   - **Cloned Or Chemically Synthesized**
4. **"Pharmaceutical" Vaccines**
   - **Peptide Derived**
   - **Orally Active**
   - **Built-in Adjuvant Effect**

***Existing Technology**

*PROTEIN THERAPEUTICS*

To change gears totally, we will now discuss how protein chemistry can help in the design of protein therapeutics. Until now, the genetic engineering approach to protein therapeutics has largely been to replicate molecules as found in nature. Using growth hormone as an example, and discounting one amino acid, what has simply been done is to replicate the natural growth hormone molecule and administer that to people who are deficient in that gene product. However, as the genetic engineering and pharmaceutical companies well know, we are now entering a phase where we seek to understand the molecules that we are going to produce well enough that we can make our production job a lot simpler. For example, we may be able to come up with more stable derivatives.

It is in this context of redesign of protein molecules that I'm going to discuss interleukin-3 and some of the work we've done on it. Interleukin-3 is one of the lymphokines. It's known by a variety of names, and in fact, when the structure of the gene was first worked out by investigators at DNAX in Palo Alto, that gene was the one coding for mast cell growth factor. IL-3 is active at many stages of hemopoiesis, and it has a unique activity on the stem cells, the pluripotential stem cells in the bone marrow, but it also acts at a number of other points, including some of the committed pathways in the development of various lineages of blood cells that are found in the circulation. Ian Clark-Lewis who spent two years working as a post-doc in my lab—he's now moved on to the Biomedical Research Centre at the University of British Columbia in Vancouver—had, as a graduate student, isolated and with our aid determined part of the structure of interleukin-3, but as panspecific hemopoietin (PSF), which turned out also to be IL-3. Figure 16 shows the complete amino acid sequence of the mature IL-3 molecule. The amino terminal was defined by our own protein sequencing efforts. The C terminal is derived from the C DNA sequencing stop codon position, which would be position 141 of the open reading frame.

Ian undertook the total chemical synthesis of this molecule two and a half years ago in my laboratory over my specific objections while I was away at a conference. It worked very well. So we've gone on and exploited that. The chemistry used on the automated synthesizer for the assembly of the protected peptide chain is shown in Figure 17. The overall steps involved in the chemical synthesis of proteins are summarized in Table IV, and the features of the total chemical synthesis approach to the study of proteins are listed in Table V.

The chemically synthesized 140 amino acid residue IL-3 protein had the complete biological activity, not just the incorporation of tritiated thymidine into factor-dependent cell lines (as shown in Figure 18, Panel A) but all the other assays for

*Figure 16*

**THE AMINO ACID SEQUENCE OF MATURE IL-3 (44).**

| | | | | |
|---|---|---|---|---|
| Ala Ser Ile Ser Gly Arg Asp Thr His Arg Leu Thr Arg Thr Leu Asn Cys Ser Ser Ile | | | | |
| 1 | 5 | 10 | 15 | 20 |
| Val Lys Glu Ile Ile Gly Lys Leu Pro Glu Pro Glu Leu Lys Thr Asp Asp Glu Gly Pro | | | | |
| | 25 | 30 | 35 | 40 |
| Ser Leu Arg Asn Lys Ser Phe Arg Arg Val Asn Leu Ser Lys Phe Val Glu Ser Gln Gly | | | | |
| | 45 | 50 | 55 | 60 |
| Glu Val Asp Pro Glu Asp Arg Tyr Val Ile Lys Ser Asn Leu Gln Lys Leu Asn Cys Cys | | | | |
| | 65 | 70 | 75 | 80 |
| Leu Pro Thr Ser Ala Asn Asp Ser Ala Leu Pro Gly Val Phe Ile Arg Asp Leu Asp Asp | | | | |
| | 85 | 90 | 95 | 100 |
| Phe Arg Lys Lys Leu Arg Phe Tyr Met Val His Leu Asn Asp Leu Glu Thr Val Leu Ala | | | | |
| | 105 | 110 | 115 | 120 |
| Ser Arg Pro Pro Gln Pro Ala Ser Gly Ser Val Ser Pro Asn Arg Gly Thr Val Glu Cys | | | | |
| | 125 | 130 | 135 | 140 |

*Figure 17*

**AUTOMATED CHEMISTRY PROTOCOL FOR THE ADDITION OF A SINGLE AMINO ACID FOR THE CHEMICAL SYNTHESIS OF *PROTEINS*.**

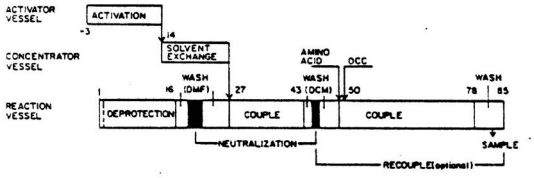



*Table IV*

**STRUCTURE-FUNCTION STUDIES OF PROTEINS BY TOTAL CHEMICAL SYNTHESIS**

1. CHAIN ASSEMBLY
2. DEPROTECTION
3. PURIFICATION
4. COVALENT STRUCTURE

THEN:

*5. ACTIVITY

*Table V*

**CURRENT FRONTIERS IN THE CHEMICAL SYNTHESIS OF PROTEINS**

- TOTAL CHEMICAL SYNTHESIS OF PROTEINS IS NOW A PRACTICAL TOOL FOR STRUCTURE-FUNCTION STUDIES: RAPID, VERSATILE (ANY SEQUENCE; UNNATURAL AMINO ACIDS; SPECIFIC LABELLING FOR NMR STRUCTURAL STUDIES)
- RATIONAL DESIGN OF PROTEINS WITH PRE-DETERMINED SPECIFIC PROPERTIES

interleukin-3. All those activities reside in the chemically synthesized sequence. We can use this total chemical synthesis to rapidly investigate various questions. For example, there are two naturally occurring forms of IL-3. Jim Ihle and his co-workers had determined the amino terminal structure of interleukin-3, and it turns out that their structure started at residue 7 of the sequence I've just shown you (Figure 17). Their molecule was 7-140, while the molecule we've been working with is 1-140, based on our own amino terminal sequence work. We chemically synthesized both molecules and, as you can see (Figure 18, Panel C) these have indistinguishable biological activities. We also used the chemical synthesis approach to carry out structure-function studies on the IL-3 molecule. Ian Clark-Lewis showed that the amino terminal 16 residues were not essential for activity, but that the 17th residue, which is the first cysteine in the sequence, is essential (Figure 18, Panel D). We then know we need cysteine 17 but we can leave out the first 16 residues if we want to make a smaller molecule.

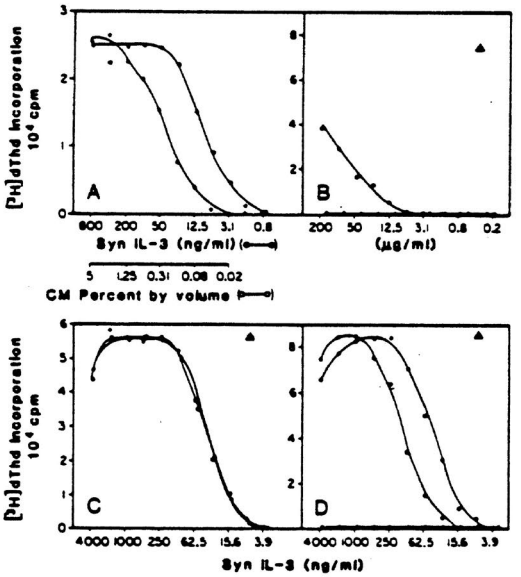
Ian then carried out a fairly typical "site-directed mutagenesis" experiment, only using total chemical synthesis of the variant forms of IL-3. He took the target sequence, which has 4 cysteines, and he replaced them all with alanines and then, knowing that 17 was essential from the experiment I've just described, he replaced, one at a time, the other three cysteines with alamines and then assayed the resulting molecules. The experimental design is shown in Figure 19. The bioassay results are shown in Table VI. The exact figures don't matter. These assays have an extraordinary dynamic range. If we plotted this in percentages and defined the [Cys 7,80, Ala 79,140] IL-3 activity then the [Cys 7,79, Ala 80,140] IL-3 would have less than 0.1% activity. The point is that cysteine 80, compared with the cysteine 79, is critical for IL-3 activity.

All the work I have just described was with the murine IL-3 molecule. Many of the companies represented here have been trying very hard to find a human IL-3 molecule, for obvious reasons. IL-3 is potentially very important as a general tonic for the immune response. There has been great difficulty in

*Figure 18*

**AUTOMATED CHEMICAL SYNTHESIS OF A PROTEIN GROWTH FACTOR FOR HEMOPOIETIC CELLS, INTERLEUKIN-3**

IAN CLARK-LEWIS, RUEDI AEBERSOLD, HERMAN ZILTENER, JOHN W. SCHRADER, LEROY E. HOOD, STEPHEN B. H. KENT




Biological activity of synthetic IL-3 and various analogues. (A) Synthetic IL-3 (1–140) (-•-) and a control preparation of tenfold concentrated WEHI-3 conditioned medium (CM) (-o-). (B) Synthetic IL-3 (1–79) (-•-) and IL-3 (80–140) (-△-). (C) Synthetic IL-3 (7–140) (-o-) and IL-3 (1–140) (-•-). (D) Synthetic IL-3 (18–140) (-△-), IL-3 (17–140) (-o-), and IL-3 (1–140) (-•-). The results shown in panels A, C, and D are from separate IL-3 syntheses. The fragments shown in panels C and D are intermediates in the IL-3 (1–140) synthesis indicated in that panel. The closed triangles in panels B, C, and D indicate the maximal activity obtained with WEHI-3 conditioned medium, a source of IL-3. The lyophilized refold synthetic peptides were dissolved in 2M guanidine-HCl and desalted on a Sephadex G-25 column equilibrated in phosphate-buffered saline. The weights were calculated from the molarity of the final preparation as determined by acid hydrolysis of the protein and amino acid analysis. The IL-3 activity was assayed with the IL-3-dependent cell line W/W.12 as described (2), except that 2 x $10^3$ cells per 10$\mu\ell$ of culture were used and the cells were treated with [³H]thymidine after 40 hours.

*Figure 19*

**IL-3: Cys REPLACEMENTS EXPERIMENTAL DESIGN**

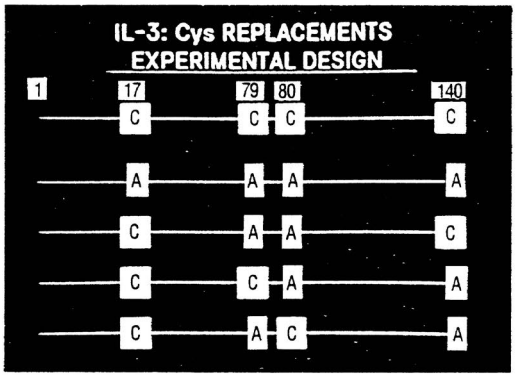



*Table VI*

**SYSTEMATIC DEVELOPMENT OF DIAGNOSTICS**

- PROTEIN MARKERS OF HUMAN DISEASE
- ANTI-PEPTIDE ANTISERA
- QUANTITATIVE 2D GELS
- PROTEIN SEQUENCE DETERMINATION

using standard molecular biology to attain this end, but in October 1986, Genetics Institute came up with a very ingenious approach and isolated a molecule which they described as human IL-3. I think they are correct in this description, but as you can see from Figure 20, these are not closely related molecules. If you line up the human molecule with the murine, where there are closed diamonds there are identical amino acids; where there are open diamonds there are conservative substitutions — (preservation of charge or hydrophilicity). You can see that even using that kind of test for relatedness, these are not very analagous molecules. These two lymphokines are much less related than other corresponding lymphokines between mouse and humans. However, one interesting thing is that in the human molecule (the comparison shown in Figure 20 is a computer generated alignment) cysteine 17 is present. The only other cysteine in the human molecule that corresponds is cysteine 80. Cysteine 79 has been replaced with a proline, and the other cysteine, in what was the C terminal of the molecule, has been deleted. There is no cysteine at the C terminal. Thus, only the single disulfide bond which we have shown to be essential for IL-3 activity is preserved in the putative human IL-3. This indicates the validity of the approach that we were taking in studying this molecule, our results also support the identification of this as the human equivalent of

*Figure 20*

**MURINE IL-3 STRUCTURE-FUNCTION: ROLE OF CYSTEINES**

*A DISULFIDE BETWEEN CYS[17] & CYS[80] IS IMPORTANT FOR ACTIVITY

*THE ONLY PAIR OF CYSTEINES IN HUMAN IL-3 IS HOMOLOGOUS:

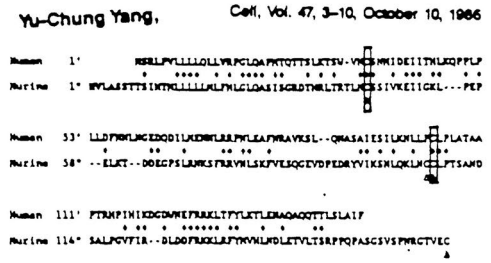



Figure 5. Comparison of the Amino Acid Sequences of the Human and Murine IL-3s

The alignment shown was generated by computer using the program of Smith and Waterman (Smith and Waterman, 1981; Waterman, 1985).

Comparison of the Amino Acid Sequences of the Human and Murine IL-3s

The alignment shown was generated by computer using the program of Smith and Waterman (Smith and Waterman, 1981; Waterman, 1985).

murine IL-3. The researchers at Genetics Institute were unaware of our results, and we were unaware of theirs.

To summarize, what I hope I've shown you is that this type of synthetic protein chemistry can help in our attempts to improve on nature. All the IL-3 work was carried out by one (talented) postdoctoral fellow in an 18 month period. What Ian Clark-Lewis has done is to take the kind of information I've just described (and he is now carrying on with his lifetime dedication to the interleukin-3 molecule in his new position in Vancouver) and he has designed much shorter molecules with only the one essential disulfide, and with other parts of the molecule deleted, but still have essentially the full activity of the IL-3 molecule. When this work is finished, in a year of two, he will then have an interesting molecule that he can then use with the chemical synthesis approach to do a lot of interesting structure-function work. In addition, if one were to produce this as a therapeutic agent, one might very seriously consider making the simpler molecule with only the one essential disulfide rather than making the parent molecule as it occurs in nature. (That is, if you were developing therapeutics for mice. If you were developing therapeutics for humans, nature has already done part of the job for you in terms of deleting one of the disulfides.)

What I want to do now is to show you a bit of work on how one uses another branch of protein chemistry, namely sequencing technology, in the development of diagnostics. As summarized in Table VI, and in no particular order, I will very quickly review protein sequencing, the use of quantitative 2-D gel technology to identify proteins in a reproducible fashion, the use of anti-peptide antisera developed from sequence information, and protein markers of human disease.

This next table (Table VII) of disease conditions studied by 2-D electrophoresis is from a book written by Allen, Saravis, and Maurer, published in 1984: Gel Electrophoresis and Isolectric Focusing of Proteins. While some of these correlations are controversial, in many cases there are characteristic protein

*Table VII*

**GEL ELECTROPHORESIS AND ISOELECTRIC FOCUSING OF PROTEINS**

R. C. Allen, C. A. Saravis, H. R. Mauer

Disease conditions studied by 2-D electrophoresis

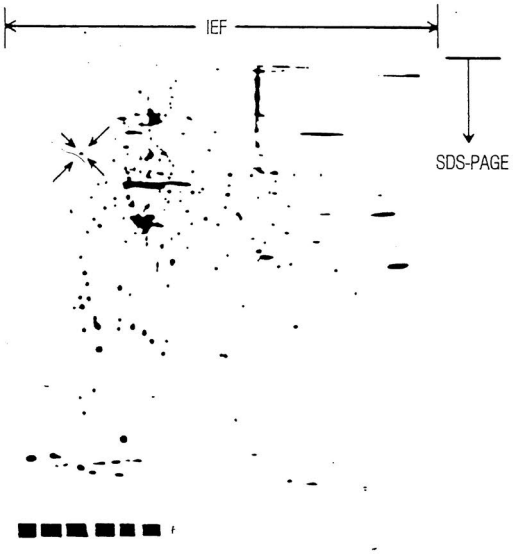
| *Disease* | *Sample* | *1st–2nd* | *Method* |
|---|---|---|---|
| Cystic fibrosis | Saliva | D/D | O'Farrell |
| Duchenne muscular dystrophy | Fibroblasts | D/D | ISO-DALT<br>Slab PAGIF<br>DALT |
| Fetal Malformation | Amniotic fluid | N/D | Slab PAGIF<br>DALT |
| Lesh-Nyhan syndrome | Lymphocytes<br>erythrocytes | D/den | O'Farrell |
| Down's syndrome | Fibroblasts | D/D | O'Farrell |
| Multiple sclerosis | Sclerotic<br>plaques | N/D | NEPHGE |
| Huntington's<br>Alzheimer's<br>Joseph's | Gliosed caudula<br>Putamen<br>Brain | | Comings |
| Myeloma proteins | Serum<br>Serum | D/D<br>N/D | ISO-DALT<br>Mod. ISO-DALT |
| Pituitary tumor | Brain | D/D | ISO-DALT |
| Intestinal carcinoma | Mucosa | D/D | ISO-DALT<br>Mod. ISO-DALT |
| Multiple myeloma | Urine | D/D | ISO-DALT |
| Neoplastic sensitive proteins | Norman and<br>transformed<br>amnion cells | D/D | NEPHGE-DALT |
| Leukaphoresis | Plasma | N/D | Agarose-<br>CZE-DALT |
| Juvenile leukemia | Mononuclear<br>cells | D/D | ISO-DALT |
| Infectious mononucleosis | Lymphocytes | D/D | ISO-DALT |
| Pancreatic adenocarcinoma | Pancreas<br>extract | D/D | Scheele |

D=denatured, N=Native protein.

markers found either in cerebro-spinal fluid, in serum or in tissues that are markers of the human disease state. And, as you are probably aware, it has been suggested over the last 10 years or so that quantitative 2-D gels actually be used as a diagnostic tool. I'd like to suggest an alternative approach.

Figure 21 shows an example of an analytical two-dimensional polyacrylamide electrophoresis gel. Depending on the exposure you take (this is an autoradiogram), there are, in fact, over a 1,000 protein species that can be separated and identified in a reproducible fashion by this technique. What I mean by identify is to put a number on the protein—not necessarily know its molecular identity—and that's the problem with the technique. Rodrigo Bravo published a book three years ago which showed how reproducible the 2D protein gel technique is. After SV-40 transformation, there is an additional protein species present. This protein is, in fact, cyclin which is now known to be a kinase sub-unit which is important for rapid cell division. This illustrates that these gels can be run in a highly reproducible fashion, so that you can in fact do a type of subtractive analysis where, particularly with the aid of computers, you can subtract the normal protein pattern from the aberrant protein pattern and be left with just the handful of proteins that have either increased or decreased dramatically in concentration.

*Figure 21*



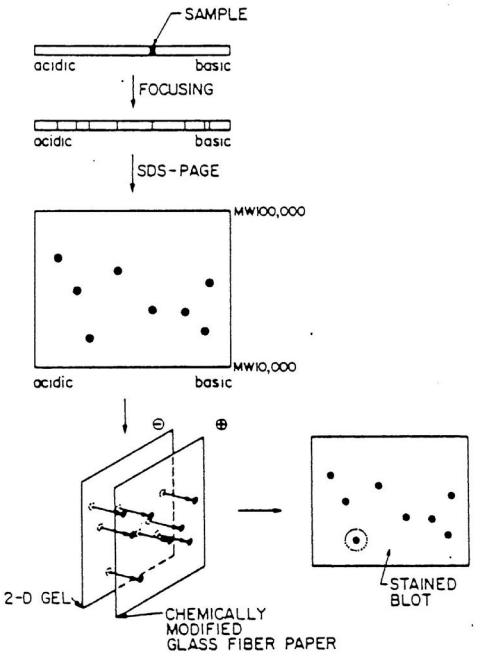

Proteins from whole-cell lysate of human fibroblasts analyzed by 2-D gel electrophoresis. Proteins were metabolically labeled with [35S]methionine and detected by fluorography. The 66kD, pI4.0 protein marked with double arrows was isolated from human lymphocytes where it is more abundant and was used in the sequence analysis described in the text (data, Figure 5).

The next question is: what are the molecular identities of the proteins? That's where things get a little bit tricky. About 30% to 40% of the proteins that are observed in *E. coli* lysates are of known molecular identity. In the case of the human fibroblasts I showed you, only a handful of proteins have been identified as molecular species even though you can separate well over a thousand. There is a general approach to the systematic identification of protein species in gels, the principle of which has been know for years and which we finally made work in our lab in the last 18 months: That is, to do amino-terminal protein sequencing of the components you are interested in directly from the two-dimensional gel.

The key to this is the use of electroblotting as a preparative method for protein sequencing. Figure 22 illustrates this. Before staining, all protein components separated in the 2D electrophoretic gel are electroblotted in the case of amino terminal sequencing onto chemically-modified glass fiber filter paper that contains positively charged moieties. In case of internal sequencing, the proteins are blotted onto nitrocellulose, as I'll mention briefly later on. You then stain the blotted proteins. In the case of the chemically modified glass fiber paper, for amino terminal sequence analysis you simply cut out the spot, put it in your sequenator, and if the protein is not blocked, you then get sequence.

To illustrate the application of this, in collaboration with John Leavitt at the Linus Pauling Institute, we examined a family of proteins that are involved in human neoplasia. These proteins migrate near 60 kDa in the acid region of the 2-D gel. We used the protein marked in Figure 21, as a model in our first attempt at using this technique. Figure 23 shows the amino terminal 10 cycles of the sequencing run resulting from the electroblotting, cutting out of that protein spot, and sequencing it. In fact, from this experiment we were able to get some 15 residues, which is pretty typical for the amounts (in few picomoles) of proteins we are dealing with. We did this for a bunch of proteins (as listed in Table VIII), and in the case of one of them we generated a synthetic peptide corresponding to the amino terminal 10 residues, and raised anti-peptide antisera in rabbits. We then went back into the 2D gel of the human cell lines that were being used for these studies, and on nitrocellulose blots we found that only the protein to which the amino terminal sequence corresponded, lit up strongly in a Western immunoblot. Generally, antisera raised against aminoterminal sequences are highly discriminating. This is illustrated in Figure 24.

*Figure 22*

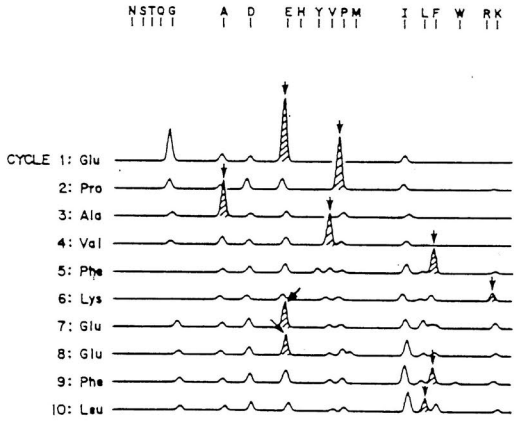



Determination of N-terminal amino acid sequence dat from proteins separated by 2-D gel electrophoresis. After the 2-D gel (IEF, SDS-PAGE) has been run, the proteins are electroblotted onto chemically modified glass filter paper and stained. The protein-containing spots revealed by staining are cut out and inserted in the cartridge of the gas-phase sequenator where the N-terminal amino acid sequence is determined by automated Edman degradation.

This is an example of how one can systematically develop an antiserum of predetermined specificity for a spot on a 2-D gel and then use that as a general way of doing immunocytochemistry; or, if you wanted to, in the developoment of diagnostics, perhaps by one of monoclonal antibodies to synthetics peptides. As summarized in Figure 25, you can use this amino terminal sequence information for the traditional things that lead to cloning and overexpression of the gene product to give you a therapeutic. Or you can go the other way via biosynthetic peptides into diagnostics.

I've mentioned that we now have a general way of overcoming amino terminal blocking of proteins which is simply to electroblot onto nitrocellulose, and then, after staining, cut out the piece of nitrocellulose that contains the protein and carry out a proteolytic digestion on the nitrocellulose (see Figure 26). The peptides are released into the supernatant and

*Figure 23*




PTH chromatograms from the first 10 steps of the Edman degradation of the PI4.0, 66kD lymphocyte protein obtained by 2-D gel electrophoresis/electroblotting (cf. Fig. 4). Elution positions of the standard PTH amino acids in the reverse phase HPLC analysis are shown. The sequence-determining PTH peaks are indicated.

*Table VIII*

**PROTEINS SEQUENCED FROM ANALYTICAL 2-D GELS.**

| *Protein* | *MW (kD)* | *pI* | *Residues Identified* | *Comments* |
|---|---|---|---|---|
| Cyclin | 35 | 4.8 | 7/8 | |
| Actin-related protein (lymphoblast) | 66 | 5.3 | 15/16 | same sequence for both |
| Actin-related protein (fibroblast) | 66 | 5.3 | 5/6 | |
| Lymphocyte polypeptide | 66 | 4.0 | 14/14 | same sequence for both |
| Fibroblast polypeptide | 66 | 4.0 | 5/6 | |
| Lymphocyte polypeptide | 70 | 5.4 | - | blocked |
| Fibroblast polypeptide | 70 | 5.4 | - | blocked |
| Lymphocyte polypeptide | 85 | 5.4 | 6/7 | |

*Figure 24*

**USE OF ANTI-PEPTIDE ANTISERA TO SELECTIVELY RECOGNIZE RELATED PROTEINS: IL-3 (1-6) VS. IL-3 (1-140) or IL-3 (7-140)**

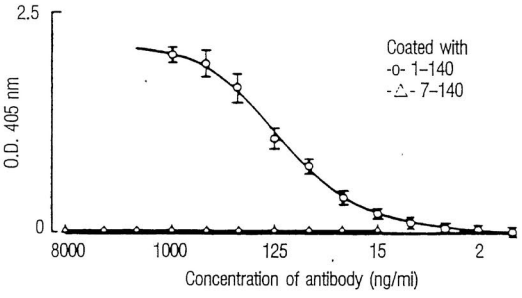
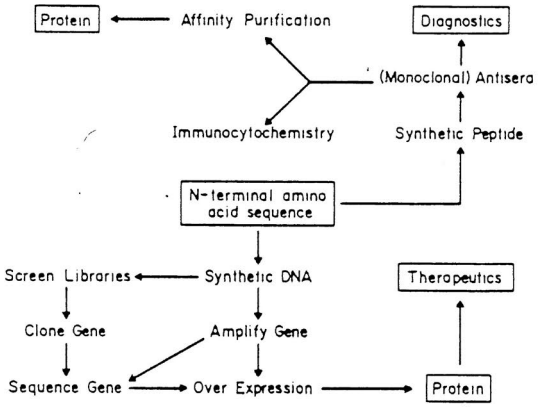


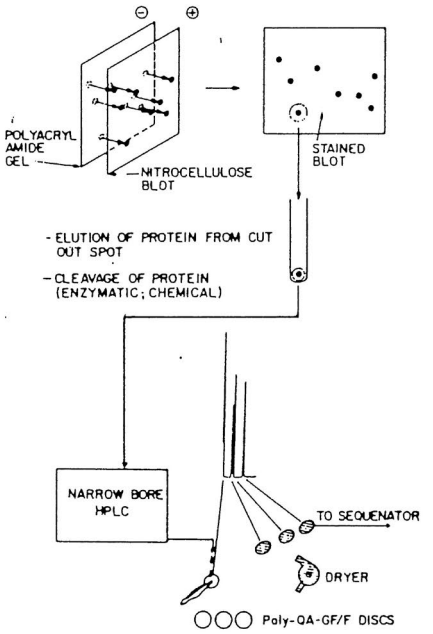

*Figure 25*




Uses of N-terminal amino acid sequence information. Acquisition of molecular structural information offers a general entry to both diagnostics (via synthetic peptides and specific antisera) and to therapeutics (via gene cloning).

*Figure 26*

**SAMPLE PREPARATION FOR INTERNAL SEQUENCING**




General approach to obtaining amino acid sequence information from N-terminal blocked proteins separated by 2-D gel electrophoresis. The proteins are electroblotted onto nitrocellulose, eluted, fragmented, and the peptides separated by reversed phase HPLC for sequence analysis. Extensive internal sequence data can be obtained in this way.

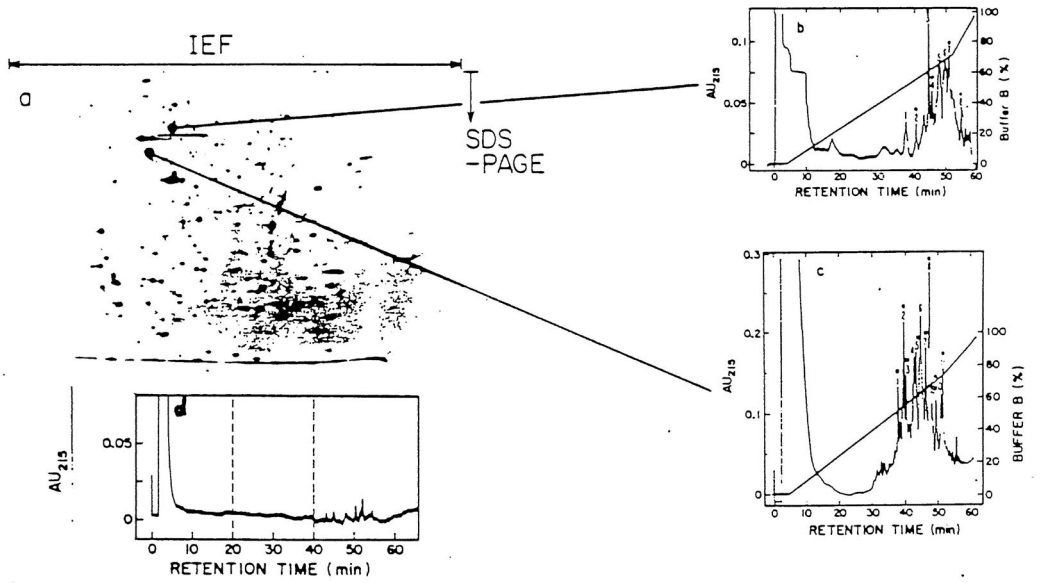
are simply separated by narow bore HPLC, as shown in Figure 28 for two proteins. The upper protein, ARP is the one that we were just discussing that we had amino terminal data from. All the peptides shown with asterisks have been successfully sequenced, in this case four or five peptides. In the case of what was believed to be the beta subunit of mitocondrial F1-ATPase, all those shown in the lower panel were separated and used to generate one hundred residues of sequence information in a week. Typical data from the Edman degradation of one of the peptides is shown in Figure 28.

The problem with general application of those techniques is one of sensitivity. We cannot sequence the amounts of protein on a single 2-D gel. We have to run up to 20 of them in order to get enough protein to be able to use existing sequencing technology, to get this kind of information (Figure 29). Although that sounds awkward, in fact it is not. It turns out that people run 10 or 20 2D gels at a time, anyway, and there are tricks that reduce the work. Ideally we would like to get the same kind of information from a single two-dimensional gel. In fact, we have new Edman chemistry and fluorescent detection sequencing working in our lab, and our initial measurements indicate that we should be able to get about a one hundred fold increase in protein sequencing sensitivity using this approach together with solid phase Edman chemistry instead of the way it's normally been done up till

*Figure 27*

**PROTEINS FROM A WHOLE-CELL LYSTATE OF THE HUMAN LYMPHOBLASTOID CELL LINE CCRF-CEM (ATCC CCL 119).**

Internal amino acid sequence analysis of proteins separated by one- or two- dimensional gel electrophoresis by *in situ* protease digestion on nitrocellulose




(a.) Two-dimensional (IEF/SDS-PAGE) separation of [$^{35}$S]Met metabolically-labeled proteins (24). Proteins used as examples are circled; (b.) HPLC map of peptides released after *in situ* tryptic digestion of the protein ARP after electroblotting onto nitrocellulose; (c.) HPLC map of peptides released after *in situ* tryptic digestion of the B-subunit of mitochondrial $F_1$-ATPase after electroblotting onto nitrocellulose. (d.) Blank digest with trypsin. The asterisk (*) indicates peptides for which sequence analysis was carried out.

now. This would be sufficient to obtain data from proteins separated in a standard analytical 2D gel.

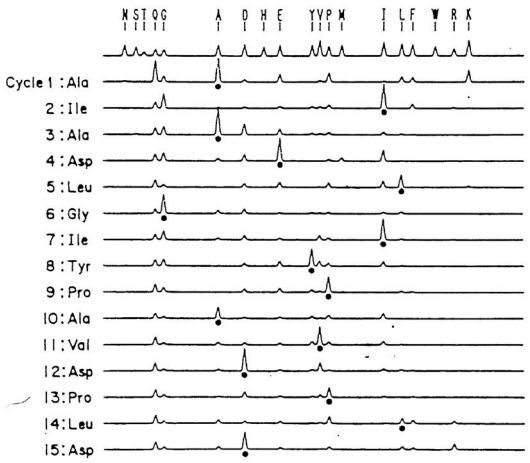
I thought I'd finish off with an example of how anti-peptide antisera can be used as diagnostics. And this time it's not on mice, it's on rats. And, of course, the reason for using these figures is that Alex Ullrich (the chairman of this session) and his co-workers have cloned the gene coding for rat, and later human insulin-like growth factor 2. The open reading frame is schematically illustrated in Figure 30. There is a leader sequence followed by a 156 amino acid open reading frame, with the four domains of insulin-like growth factor 2, giving rise to a 67 residue protein product. Then there is this last E-domain which is conserved in different species and, therefore, presumably does something important. With Dan Straus at U.C. Riverside we showed that a peptide corresponding to the C-terminal 40 residues of this is actually found in the serum of rat—the last 40 residues, 117 to 156. We identified this product by protein sequence analysis and then by total chemical synthesis and we used the synthetic 40 residue peptide to make an anti-peptide antiserum, which Dan Straus used to set up a very nice and very sensitive immunoassay. It was very easy to detect very small amounts of the naturally occurring peptide, using the anti-peptide antisera. There is no cross reactivity with other related molecules or molecules of similar activities.

If we then use this as a radioimmunoassay, we ask where does this E peptide occur naturally, and at what levels? Figure 32 shows the result of one such experiment with rats. You can see that there are low levels of E peptide in adult rats, but in pups immediately after birth there are very high levels, almost 100 fold higher. This is an example of the use of anti-peptide antiserum as the basis for a radioimmunoassay and its application *in vivo*.

To summarize, I've shown how protein chemistry can be

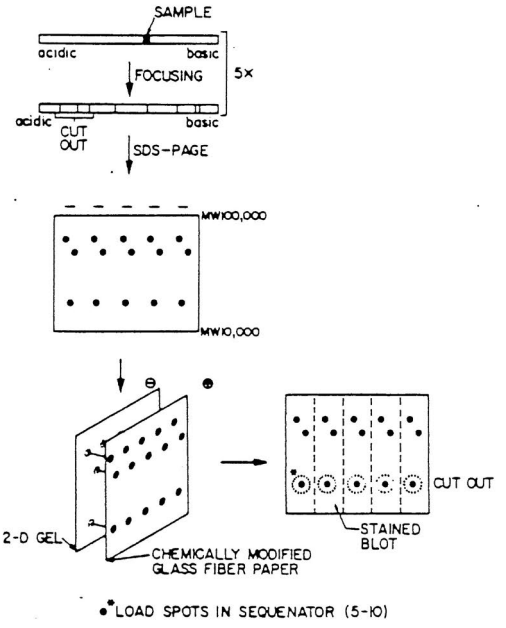
*Figure 28*

**HPLC ANALYSES (18) OF THE PHENYLTHIOHYDANTOIN (PTH) AMINO ACIDS RELEASED FROM THE FIRST FIFTEEN CYCLES OF EDMAN DEGRADATION OF $F_1$ ATPASE PEPTIDE -8 (Fig. 2c). A SEPARATION OF PTH STANDARDS (10 PMOL EACH) IS SHOWN, WITH AMINO ACIDS IDENTIFIED BY THE SINGLE LETTER CODE.**




Chromatographic data was digitally acquired and the displayed chromatograms have been mathematically filtered to remove high frequency noise and baseline shifts (19). PTH signals defining the major sequence present are indicated by dots (•) beneath each peak. A second minor sequence is present.

*Figure 29*




A simple approach for increasing the amount of protein isolated from 2-D gels for sequence analysis. Multiple (5–10) first (IEF) dimensions are run, the pH range of interest is cut out and the multiple gel pieces loaded across the top of a single second (SDS-PAGE) dimension. After the gel has been run, electroblotted, and stained, the 5–10 copies of each spot are cut out and used for protein sequence analysis.

used to help design proteins as products, and to understand them better before we go into the design phase; and also what may well be a general approach to the development of diagnostics for some human diseases. There are a couple of points I thought I might make. The first is that there's been a lot of talk about vaccines today, and most of the people sitting here are from industry and well aware of the problem with vaccine development. It seems to me that with two of the important diseases that we've discussed today, namely AIDS and hepatitis B, this is a particularly acute problem. For what is happening is that the Third World is acting as a reservoir for these diseases and eventually reinfecting the population in our own country and others. We have to address the origins of the diseases. The problem, of course, is economics. The fact is that you are not going to use a $10 a dose vaccine to immunize the populations of Asia and Africa to eradicate hepatitis B (the current vaccine is $100 for the series). So I don't have an answer.

The other thing I should point out is the reason there has been a great deal of emphasis on vaccines is that they are probably *the* single way that you can have the biggest impact on public health.

Finally, I hope it is clear that protein chemistry does, in fact, have a substantial contribution to make to the development of products in biotechnology.

## QUESTIONS AND ANSWERS

Question: Can you comment or hypothesize on the origin of the truncated IL-3 species that had 6 amino acids missing?

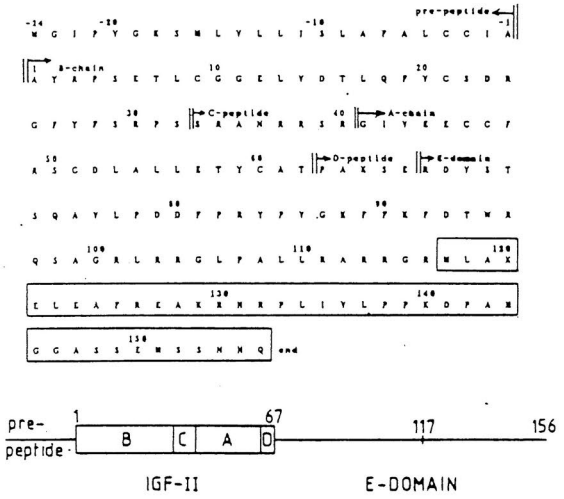
Response: That (7-140)IL-3 molecule was isolated from conditioned cell media, and we believe it is a proteolytic cleavage artifact. Cleavage occurs between residues Arg 6 to Asp 7, a likely proteolysis site. We now have a recent paper, I believe in the February issue of *General Immunology*, where we develop anti-peptide antisera against the amino terminal 6

*Figure 30*

**E-DOMAIN PEPTIDE OF RAT PROINSULIN-LIKE GROWTH FACTOR-II: VALIDATION OF A RADIOIMMUNOASSAY AND MEASUREMENT IN CULTURE MEDIUM AND RAT SERUM***

VINCENT W. HYLKA, STEPHEN B. H. KENT, and DANIEL S. STRAUS

*Division of Biomedical Sciences and Department of Biology, University of California (V.W.H., D.S.S.), Riverside, California 92521-0121; and the Division of Biology, California Institute of Technology (S.B.H.K.), Pasadena, California 91125*



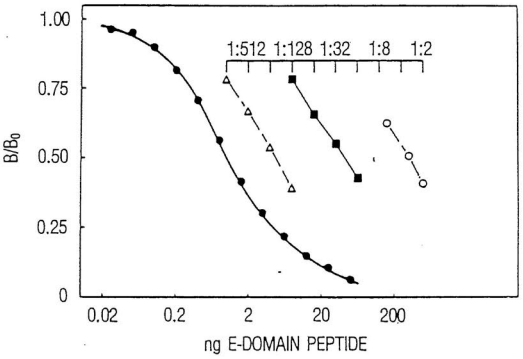

Rat prepro-IGF-II, as predicted from cDNA sequence analysis (11, 12). The amino acids in the *boxed area* represent the sequence of the synthetic E-domain peptide, pro-rIGF-II-(117–156). The Phe$^{117}$ analog is a 40-amino acid peptide with Phe substituted for Met$^{117}$. The Tyr$^{116}$ analog is a 41-amino acid peptide with Tyr substituted for Arg at position 116. A schematic of the spatial relationships of IGF-II and the E-domain is shown below the sequence data. The numbers in the schematic represent the prepro-IGF-II residue position.

residues of the 1-140 sequence. As was shown in Figure 26, these are extremely highly discriminating. They recognize only the full length sequence, as you would expect, and they also show very high affinity for the native molecule. We use this as a reagent to investigate this exact point. We found that all the IL-3 that is released from antigen or mitogen-activated T-cells is the full length (1 to 140)IL-3 and you only find the (6- 140) IL-3 in conditioned cell media. The implication is that it is a proteolysis artifact.

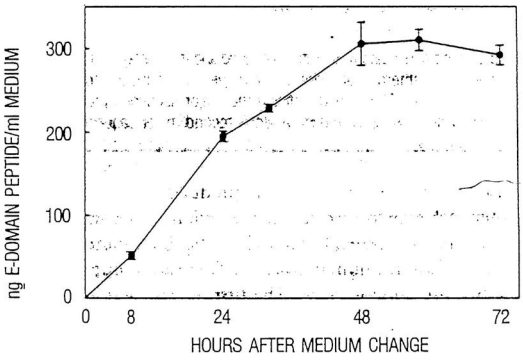
Question: Let me ask you a question regarding your protein sequencing strategy from 2-D gels. How many proteins are amino-terminally blocked when you use this approach?

*Figure 31*

**E-DOMAIN PEPTIDE OF RAT PRO-IGF-II**




RIA standard curve and parallelism of E-domain peptide in neonatal rat serum, adult rat serum, and serum-free medium conditioned by BRL-3A rat liver cells. The dilution scale for medium or serum is shown at the *top* of the figure. •, E-Domain peptide standard curve; △, 5-day old neonatal rat serum dilutions; ○, adult rat serum dilutions; ■, dilutions of BRL-3A-conditioned medium. $B/B_0$, Bound to 100% bound ratio.




Measurement of E-domain peptide secreted by BRL-3A rat liver cells into serum-free medium as a function of time. Values represent the mean ± SE of five individual flasks of cells. All values were determined within a single assay. Similar results were obtained in a second experiment performed on a different day with a different batch of cells.

Response: Twenty to fifty percent are amino terminally blocked. This figure is based on attempting to sequence 20 to 30 proteins so far, both amino terminally and internally. One of the things we are now undertaking, is the essentially almost mindless exercise of sequencing the 100 most abundant spots in these 2-D gels. I think we'll learn a lot. It's not

the kind of research I normally approve—I call it "grass gathering"—but in this case it may be useful.

*Acknowledgements*
The hepatitis B virus research was performed by Karen Parker and was part of a continuing collaboration with Dr. Robert Neurath (The New York Blood Center). The IL-3 work was carried out by Dr. Ian Clark-Lewis in my laboratory at the California Institute of Technology. The protein sequence analysis of 2D gels was carried out by Dr. Ruedi Abersold in a program jointly directed by Stephen Kent and Leroy Hood.