Automatic Sequence Design of Major Histocompatibility Complex Class I Binding Peptides Impairing CD8+ T Cell Recognition*§

Koji Ogata‡, Alfonso Jaramillo‡, William Cohen§, Jean-Paul Briand¶, Francine Connan§, Jeannine Choppin§, Sylviane Muller§, and Shoshana J. Wodak‡‡

From the ‡Service de Conformation de Macromolécules Biologiques et Bioinformatique, CP263, Centre de Biologie Structurale et Bioinformatique, Université Libre de Bruxelles, Blvd. du Triomphe, 1050 Brussels, Belgium, the §Département d’Immunologie, INSERM U567-CNRS UMR 8104-Université Paris V René Descartes, Institut Cochin IFR 116, 27 Rue du Faubourg Saint-Jacques, 75014 Paris, and the ¶Service de Conformation de Macromolécules Biologiques et Bioinformatique, CP263, Centre de Biologie Structurale et Bioinformatique, Université Libre de Bruxelles, Blvd. du Triomphe, 1050 Brussels, Belgium

Received for publication, July 9, 2002, and in revised form, October 13, 2002

Published, JBC Papers in Press, October 30, 2002, DOI 10.1074/jbc.M206853200

This paper is available on line at http://www.jbc.org

© 2003 by The American Society for Biochemistry and Molecular Biology, Inc.

Printed in U.S.A.

An automatic protein design procedure was used to compute amino acid sequences of peptides likely to bind the HLA-A2 major histocompatibility complex (MHC) class I allele. The only information used by the procedure are a structural template, a rotamer library, and a well established classical empirical force field. The calculations are performed on six different templates from x-ray structures of HLA-A0201-peptide complexes. Each template consists of the bound peptide backbone and the full atomic coordinates of the MHC protein. Sequences within 2 kcal/mol of the minimum energy sequence are computed for each template, and the sequences from the templates are combined and ranked by their energies. The five lowest energy sequences of five other low energy sequences are ranked on the basis of their similarity to peptides known to bind the same MHC allele are chemically synthesized and tested for their ability to bind and form stable complexes with the HLA-A2 molecule. The most efficient binders are also tested for inhibition of the T cell receptor recognition of two known CD8+ T effectors. Results show that all 10 peptides bind the expected MHC protein. The six strongest binders also form stable HLA-A2-peptide complexes, albeit to varying degrees, and three peptides display significant inhibition of CD8+ T cell recognition. These results are rationalized in light of our knowledge of the three-dimensional structures of the HLA-A2-peptide and HLA-A2-peptide-T cell receptor complexes.

Major histocompatibility complex (MHC)† class I molecules are cell surface glycoproteins, which consist of a highly poly-

* This work was carried out as part of a project entitled “A Multidisciplinary Approach to the Development of Epitope-based Vaccines” and was supported by European Communities Grant BIO4CT980294, the Fonds National de la Recherche Scientifique Belge, and the Action de Recherches Concertées de la Communauté Française de Belgique. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† The abbreviations used are: MHC, major histocompatibility complex; TCR, T cell receptor; APLs, altered peptide ligands; β2m, β2-microglobulin; PDB, Protein Data Bank; CHAPS, 3-(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate; FMOC, N-(9-fluorenyl)-methoxycarbonyl; RP-HPLC, reversed-phase high pressure liquid chromatography; Ab, antibodies; PBMCs, peripheral blood mononuclear cells; IL, interleukin; IFN, interferon; PBS, phosphate-buffered saline; r.m.s., root mean square; CTL, CD8+ T cells.
other T cell subsets (12–16). Design of novel peptides with both high HLA binding affinity and capable of targeting specific HLA alleles can therefore also be used for modulating the immune system.

Many studies devoted to these goals have used theoretical approaches to predict peptide sequences with the required affinity and selectivity for specific MHC molecules (for review see Ref. 17). Most of these approaches relied on sequence information alone (18). By and large, they trained Neural Networks (19) or Hidden Markov Models (20) on peptide sequences shown experimentally to bind, or not to bind, a specific MHC allele, and then used these models to predict new peptide sequences likely to specifically recognize the same protein. The majority of these studies were done on class I molecules, where peptide recognition is more specific and better understood than in the class II molecules (19, 21).

In this paper we report an alternative approach to the prediction of peptide sequences that bind to MHC class I molecules. This approach uses an automatic procedure, implemented in the software DESIGNER (22) for selecting amino acid sequences that are compatible with a given peptide backbone conformation. This conformation is considered in the context of the three-dimensional structure of a specific MHC class I allele, the human HLA-A0201 allele, to which the peptide is bound.

DESIGNER has been developed as a general procedure for selecting families of amino acid sequence likely to fold into the three-dimensional structure defined by a given template (22, 23). Selected sequences are those that minimize a fitness function, which represents the protein free energy of folding. This function relies on basic physical chemical principles that underlie molecular interactions and protein stability. It combines the all atom force field of CHARMM (24, 25) and an implicit hydration term that depends linearly on the solvent-accessible surface area of the solute (26). The electrostatic term is computed using a dielectric constant of 8 and a switching function operating at a distance of 6 Å. Side chain conformations are modeled using the backbone dependent rotamer library of Dunbrack and Karplus (33). At each position, all natural amino acids are considered, except prolines. The free energy of the reference state is computed as the sum of the free energies of isolated amino acids, using exactly the same force field.

To predict peptide sequences that bind to the HLA-A0201 molecule, DESIGNER was run on six different templates derived from representative x-ray structures of HLA-A0201-peptide complexes deposited in the PDB (31), as shown in Fig. 1. For each of the six templates, DESIGNER produced all the sequences within 2 kcal/mol of the minimum energy sequence, and the sequences computed for all the six complexes were combined and ranked in order of increasing energy. In addition, the computed low energy sequences were re-ranked on the basis of their similarity to sequences of peptides retrieved from the MHCPEP data base (32) and known to bind the same MHC allele.

The five best ranking peptide sequences produced by the pure free-energy ranking and five other sequences obtained by the combined energy and similarity ranking were retained for experimental analysis. The 10 peptides were chemically synthesized and tested for their ability to bind HLA-A2 molecules and to form stable HLA-A2-peptide complexes. In addition, the most efficient binders were tested for their capacity to inhibit the recognition of two known CD8+ T effectors.

The remarkable results reported here are that all 10 peptides were found to actually bind the HLA-A2 molecules. Six peptides displayed 50–113% of the binding activity measured with the natural peptide used as control, as well as significant MHC stabilization. Furthermore, of the six most active peptides three displayed significant inhibition of the CD8+ T responses. In the following these findings are described in detail. They are rationalized in light of our current understanding of the factors that determine MHC-peptide interaction and TCR recognition, and their implications for the design of peptides that modulate the MHC-mediated immune response are discussed.

MATERIALS AND METHODS

Computational Procedures

Selection of Template Structures—The HLA-A0201 allele type chosen for this work has the largest number of structures of MHC-peptide complexes deposited in the Protein Data Bank (PDB) (31). The April 2000 release of the PDB contained a total of 13 complexes with 9-residue peptides with different sequences and conformations. To ensure adequate exploration of sequence space for the bound peptide by the sequence design procedure (see below), this procedure was applied to several of these complexes, used as structural templates. To select the templates, the peptide backbones from all 13 complexes were superimposed, and whenever two peptide backbones were within 1 Å r.m.s. deviation of one another, they were considered as too similar, and one of the complexes (with the lowest resolution structure) was discarded. This yielded six MHC-peptide complexes with the following PDB codes: 1AKJ, 1BD2, 1B0G, 1HHG, 1HHK, and 1HHI, which were used as template for the sequence calculations.

Automatic Sequence Design of MHC-binding Peptides—The peptide sequences are predicted using the procedure implemented in the software DESIGNER, described previously (22). This procedure selects amino acid sequences compatible with a given structural template. It has two main components as follows: the fitness function, which measures the fitness of a given sequence for the structure at hand, and the optimization procedure, which selects highly scoring sequences from a very large number of possibilities.

The fitness function is a quantity akin to the folding free energy. This quantity is computed as the difference between the free energies of the protein native folded state and a reference state used as a model for the protein unfolded state.

The free energy of the folded state comprises an interaction energy with terms computed using the standard all atoms molecular mechanical force field of CHARMM (24, 25), and an implicit hydration term that depends linearly on the solvent-accessible surface area of the solute (26). The electrostatic term is computed using a dielectric constant of 8 and a switching function operating at a distance of 6–7 Å. Side chain conformations are modeled using the backbone-dependent rotamer library of Dunbrack and Karplus (33). At each position, all natural amino acids are considered, except prolines. The free energy of the reference state is computed as the sum of the free energies of isolated amino acids, using exactly the same force field.

To select amino acid sequences with lowest free energies, we used both the Dead-End elimination and a heuristic procedure with 250,000 iterations (22).

Peptide sequences were selected by running DESIGNER on each of the six representative MHC-peptide complexes of the HLA-A0201 allele. For each complex, the bound peptide was stripped of its side chains. The bare peptide backbone and both the backbone and side chains of the MHC molecule were kept fixed at their crystallographic positions, constituting the fixed structural template, in whose environment the peptide sequence selection was performed. For each of the six complexes, DESIGNER produced all the sequences within a 2 kcal/mol of the minimum energy sequence, and the sequences computed for all the six complexes were combined and ranked in order of increasing energy.

Whenever the bound peptide used as template contained a Pro residue, this residue was re-modeled as an Ala residue, by removing the Cβ atom; the backbone was not re-modeled.

Use of Profiles of MHC Binding Peptides to Re-rank Peptide Sequences Selected by DESIGNER—The above-described sequence selection procedure should in principle produce a family of sequences corresponding to peptides that form stable complexes with the considered MHC allele. By having ranked these sequences in order of increasing free energy, those most likely to yield stable complexes are expected to appear at the top of the list. This is however not guaranteed, given the inaccuracies of the sequence selection procedure. It therefore seemed useful to apply in addition another selection criterion, based on sequences of peptides that are known to bind the same MHC allele. Such a criterion was obtained by scoring the predicted sequence against the profile derived from sequences of peptides known to bind to the HLA-A0201 allele.

This profile was derived from sequences stored in the MHCPEP data base (32). Because this data base contains very similar peptide sequences, the sequences were first clustered into groups with more than
50% sequence identity. For each group the occurrence of each of the 20 natural amino acids was recorded at each sequence position (9 in all) as a binary event (a given amino acid either occurs at a given position or it does not). The number of times an amino acid occurred at a given sequence position in all groups was then determined and used to derive the position-specific amino acid frequency matrix, or probability p(a,k), of having the amino acid a, at position k along the sequence.

The similarity score Sim, of a peptide sequence against this profile was computed as shown in Equation 1,

$$Sim = \sum_{k} w_k \cdot p(a, k) \quad (Eq. 1)$$

where $w_k$ are position-specific weights whose values were set to 1 for all k.

All the peptides selected by DESIGNER as having a free energy of 2 kcal/mol above the minimum were then re-ranked in order of decreasing Sim value. Those with the highest Sim values were considered as most resembling the sequences of peptides known to bind the considered MHC allele, and hence as most likely to yield a stable complex.

**Experimental Procedures**

**Peptide Synthesis**

Peptides were synthesized by Fmoc chemistry with the stepwise solid-phase methodology using a multichannel peptide synthesizer (34). Protected amino acids were coupled by *in situ* activation with (benzotriazol-1-yloxy)tris(dimethylamino)phosphonium hexafluorophosphoate, and N-Fmoc de-protection was performed as described previously (34). Side chain protection and cleavage of peptides from the solid support was performed by treatment with reagent K (82.5% trifluoroacetic acid, 5% phenol, 5% water, 5% thioanisole, 2.5% 1,2-ethanediol) for 2h 30 min at 20 °C (35). Peptides were purified by reverse-phase HPLC (RP-HPLC) using a PerkinElmer Life Sciences preparative HPLC system on an Aquapore ODS 20-μm column (100 × 10 mm). Elution was performed with a linear gradient of aqueous 0.1% trifluoroacetic acid (A) and 80% acetonitrile containing 0.1% trifluoroacetic acid (B) at a flow rate of 6 ml/min with UV detection at 220 nm.

Analytical RP-HPLC was run on a Beckman Instruments (Gagny, France) with a Nucleosil C18 5-μm column (150 × 4.6 mm) using a linear gradient of 0.1% trifluoroacetic acid in water and acetonitrile containing 0.1% trifluoroacetic acid at a flow rate of 1.2 ml/min. Matrix-assisted laser desorption and ionization time-of-flight spectra were obtained on a Protein TOF™ mass spectrometer (Bruker, Wissembourg, France).

**Preparation of HLA-A2 Molecules**

HLA-A2 molecules were purified as described previously (36) from Epstein-Barr virus-transformed B cell lines by affinity columns coupled to BB7.2 antibodies (Ab) directed against HLA-A2 molecules (HB82, ATCC, Manassas, VA) and frozen at –80 °C. HLA molecules were denatured in phosphate-buffered saline (PBS) containing 12.5 mM NaOH (pH 11.7) and 1.5 M urea for 1 h at 4 °C. The heavy chains and β2-microglobulin (β2m) were separated from endogenous peptides on a Sephadex G-25 column (PD10, Amersham Biosciences) equilibrated in PBS containing 0.05% Tween 20, 10 mg/ml bovine serum albumin, 1 mM phenylmethylsulfonyl fluoride, 10 μg/ml trypsin inhibitor, 3 μM sodium azide (PBS-Tw). Then 2 μg/ml exogenous β2m (Sigma) and 6 mM CHAPS were added just before addition of exogenous peptide.

**Detection of Peptide/HLA Interactions**

**Direct Binding Test**—The conditions of the test were described previously (36). Briefly, 50 μg of denatured HLA molecules in 2.5 ml of PBS containing all the additives described above were added in aliquots containing 1 μg of HLA and incubated with exogenous peptide at 10⁻⁵ to 10⁻⁸ M, in Eppendorf microtubes for 2h at room temperature and then for 24 h at 4 °C. Each aliquot was further divided into 2 wells of a microtiter plate (100-μl well, MaxiSorp Nunc) coated with anti-HLA-A2 BB7.2 Ab (10 μg/ml in PBS) and incubated for 20 h at 4 °C. Correctly folded HLA complexes were assessed using anti-β2m monoclonal Ab M28 coupled to alkaline phosphatase and 4-methylumbelliferyl phosphate as a substrate (Sigma). Fluorescence was measured at 355/460 nm (VICTOR™, Wallac, Evry, France). The positive control was a viral CHS tested by HLA-A2. M58-66 (GILGFVFTL) from matrix of influenza virus, and the negative control was an epitope presented by HLA-B27, NP.383–391 (SRYWAIRTR) from the nucleoprotein of influenza virus.

**Stability of Peptide-HLA Complexes**—HLA denaturation and renaturation with 10⁻⁶ M exogenous peptide were performed as described above. After an overnight incubation, unbound peptides were removed by centrifugation on NANOSEP™ 10K (Pall Filtron, Northborough, MA). Samples were dialyzed in PBS-Tw and aliquoted into Eppendorf microtubes for further incubations at 37 °C for 1, 3, 5, 24, or 48 h. One aliquot (time 0) was tested immediately to measure the maximum number of complexes. The incubations with anti-HLA-A2 BB7.2 Ab and anti-β2m M28 Ab, respectively, were performed in microplates at 37 °C for 1 h. Final detection was performed as described above.

**Inhibition of T Cell Responses**

*Generation of Effector T cells*—Human T cell effectors were generated as described previously (16) after *in vitro* stimulation of peripheral blood mononuclear cells (PBMCs) from HLA-A2-possitive donors with synthetic peptides. Unfractionated PBMCs were seeded in 24-well plates with 1 μM of tetanus toxoid and 1 μM peptide M58-66 or MART (Leu-27) 26-35 (ELAGIGILTV). Interleukins IL-7 and IL-2 were added as reported previously. Anti-M.58-66 T cell effectors were obtained after 3 weekly stimulations and anti-MART (Leu-27)-(26-35) T cell effectors after 6 stimulations.

*T Cell Recognition of HLA-A2-Peptide Complexes*—This was assessed using an Enzyme-linked Immunoassay (ELISpot) assay detecting secretion of IFN-γ (36). 96-well nitrocellulose plates (Millipore, Bedford, MA) were coated with 2 μg/ml mouse anti-human IFN-γ monoclonal (number 1598-00 Genzyme, Russelheim, Germany). PBMC, either freshly isolated or thawed, were cultured overnight in complete medium and plated in triplicate at serial dilutions (3 × 10⁻¹⁰ M cells per well).

Appropriate stimuli were then added, and the plates were incubated for 20 h at 37 °C in 5% CO₂. After washing, the cells were incubated with 100 μl of rabbit polyclonal anti-human IFN-γ antibody diluted 1:250 (IP500, Genzyme), then with a biotinylated anti-rabbit immunoglobulin G diluted 1:500 (Roche Molecular Biochemicals), and finally with alkaline phosphatase-labeled extravidin (Sigma). Spots were developed by adding chromogenic alkaline phosphatase substrate (Bio-Rad), and colored spots were counted in a stereomicroscope. A result was considered as significant when the numbers of spots were at least twice the background (the value given by negative peptides) and were proportional to the numbers of plated cells. Frequencies of IFN-γ spot-forming cells were calculated. Positive controls for interferon detection consisted of 6 wells containing 300–1000 cells stimulated with 50 ng/ml phorbol myristate acetate and 500 ng/ml ionomycin. This strong mitogenic stimulus verified that freezing and thawing did not introduce artifacts and was an indirect check of overall T cell viability. Negative controls for immune recognition consisted of epitopes derived from various viruses (for example, peptide Tax 11–19 from HTLV-I, which associates with HLA-A2) and never elicited a significant response compared with PBMC incubated in medium alone.

*Inhibition Assay*—Presenting T cells were sensitized with 5 × 10⁻⁸ M peptide MART (Leu-27)-(26-35) and 10⁻⁶ M peptide M58-66, respectively. Peptides being tested as inhibitors were simultaneously added at molar concentrations varying from 0.1 to 1000 M. Appropriate stimuli were then added, and the plates were incubated for 1 h. Final detection was performed as described above.

**RESULTS**

**Prediction of Peptide Sequences That Bind MHC Class I Molecules**—The sequence design procedure DESIGNER was run using as templates the three-dimensional structures of each of the six HLA-A201-peptide complexes listed in Table I and illustrated in Fig. 1. This yielded between 98 and 702 different sequences for individual templates, whose energies were within a 2 kcal/mol window of the minimum energy sequence computed for each template. Combining the results from all six templates the number of low energy peptide sequences totaled 1430.

From these sequences, 10 were selected as the most likely candidates for HLA-A2 binding. These are numbered 1–10 and are listed in Table II. Sequences 1–5 correspond to the highest ranking (or lowest energy) sequences from among all the 1430 designed sequences. Sequences 6–10 are those from among the same 1430 peptides, but which display the highest sequence similarity score s, relative to peptide sequences known to bind to the HLA-A0201 allele. This similarity score was computed using a sequence profile derived from 903 HLA-A0201-binding,
particular, all the sequences feature a Leu residue in P2 but whereas the lowest energy peptides have (Leu, Ile, Met, or Val in position 2 of the peptide (P2), and Val, have one of the anchor residues required for HLA-A2 binding the 1B0G entry. Interestingly, all five lowest energy sequences and three of the two sequences with high contributions most to the interactions with the TCR (5).

Table I
Template structures of MHC-peptide complexes used in the prediction

| PDB code | Peptide sequences | Peptides source | ∆ASA | No. H-bonds | No. predictive sequences |
|----------|------------------|----------------|-------|-------------|-------------------------|
| 1AKJ     | ILKEPVHGV        | HIV-1 RT       | 1089.4| 4           | 702                     |
| 1B0G     | ALWGFFFPVL       | Human peptide P1049 | 1099.3| 1           | 140                     |
| 1HHG     | TLTSCTSV         | HIV-1 gap120   | 984.9 | 4           | 238                     |
| 1HHI     | GILGFVFTL        | Influenza A matrix | 1087.3 | 1   | 105                   |
| 1HHK     | LLFGYPVYY        | HTLV-1 Tax     | 1067.3| 0           | 98                      |

* ∆ASA: Difference of the solvent-accessible surface area between the bound and unbound peptide. The solvent-accessible surface area was computed using the software Access, which implements the procedure of Ref. 41.

9-residue-long peptide sequences stored in the last release of the MHCPEP data base (32), as described under “Materials and Methods.”

Table II lists for all the selected sequences, the energy computed by DESIGNER, the value of their similarity score Sim, and the maximum identity score with the HLA-A201-bound peptide sequences in the MHCPEP data base. As expected, the first five peptides have energies ranging from −55.391 to −54.811 that are about 1.8–6.6 kcal/mol lower than the last 5 peptides. In contrast, the latter all have s values exceeding 1.2, whereas the lowest energy peptides have s values below 1. Interestingly, all five lowest energy sequences and three of the five sequences with highest Sim values have been obtained using the template from the 1HHK RSCB-PDB entry. The remaining two sequences with high Sim values have been designed using the 1B0G entry.

Inspection of the 10 sequences clearly reveals that they all have one of the anchor residues required for HLA-A2 binding (Leu, Ile, Met, or Val in position 2 of the peptide (P2), and Val, Leu, Ile, Ala, or Met at the peptide C terminus (PÎ)). In particular, all the sequences feature a Leu residue in P2 but either Leu or Val at PÎ. This difference in anchor residue variability is most likely related to the difference in conformational variability of the corresponding binding pockets in the MHC molecule. We could verify that the r.m.s. deviations of MHC residues lining the Ï pocket in the six considered templates are on average larger (0.8–1.3 Å) than those of the B pocket (0.2–0.6 Å).

It is furthermore noteworthy that the first five lowest energy peptides have very similar sequences, with only the central positions 4 and 5 and to a lesser extent position 1 (which is aromatic, but either Phe or Tyr) displaying some variability. In the last five highest scoring peptides the sequence variability is more pronounced and spreads over positions 3–8. All the predicted peptide sequences in Table II except for one (FLLRDRIFV, peptide number 3) display at most 55% sequence identity to known MHC binders, corresponding to a moderate statistical significance (p value of 6.1e−4).

To illustrate the variability of the designed sequences and how it compares with that of their natural counterparts from the MHCPEP data base, Table III lists the position-specific amino acid frequency matrices, or profiles, of both sequence ensembles. Inspection of these profiles shows clearly that the natural sequences of the HLA-A2 binders are significantly more diverse than the designed sequences, in all positions except for the anchoring residues, which are quite conserved in both sequence groups. The diversity of the natural sequences in those positions is too large for deriving a meaningful consensus sequence from which sequences of HLA-A2 binders can be predicted. The position-specific entropy computed from the corresponding amino acid frequencies, as detailed in the legend of Table III and listed in the last row of this Table, reflects well these properties. In the MHCPEP sequences this entropy is quite large (near its maximum value of 2.996 = −log (1/20)) in all positions save those of the two anchor residues. In the designed sequences, the position-specific entropy is lower for all positions, yet that of the middle residues (positions 5–7) is nearly twice as high as the entropy of the two anchor residues.

Close inspection of the two profiles in Table III furthermore reveals that the designed sequences reproduce several of the trends observed in the natural sequences. For instance, Asp occurs most often in position 4, in both the natural and designed peptides; Glu occurs frequently in positions 4, 5, and 8, whereas Phe occurs most often in position 1.

We thus see that although our sequence design procedure yields predictions that could not have been deduced readily from the analysis of the natural peptide sequences, these predictions nonetheless reproduce key features of these sequences.

Inherent in our procedure for predicting peptide sequences is the fact that it also predicts their most likely three-dimensional conformation in the environment of the MHC molecule. The conformations of the lowest energy peptides predicted using the 1HHK and 1B0G templates, respectively, are displayed in...
Automatic Design of MHC Class I Binding Peptides

The sequences numbered 1–5 are the 5 lowest energy peptide sequences computed by the protein design procedure DESIGNER. The sequences numbered 6–10 are those among the lowest energy sequences selected by DESIGNER that display the highest similarity score.

**TABLE II**

Highest scoring sequences of peptides likely to bind the HLA-A2 molecule

| No. | Sequences | Template | Energy | Sim | Maximum identity | ΔASA | No. H-bond |
|-----|-----------|----------|--------|-----|------------------|------|-----------|
| 1   | FLLQWRIFV | 1HHK     | -55,391| 0.995| 55.6             | 1084.3| 0         |
| 2   | FLLQDRIFV | 1HHK     | -55,292| 0.982| 55.8             | 1072.4| 2         |
| 3   | FLLRDHIFV | 1HHK     | -55,268| 0.982| 44.4             | 1071.4| 2         |
| 4   | YLLLQRFIV | 1HHK     | -54,910| 0.956| 55.8             | 1083.2| 1         |
| 5   | YLLLQDIFV | 1HHK     | -54,811| 0.944| 55.8             | 1071.7| 3         |
| 6   | FLLQWQIKL | 1HHK     | -53,583| 1.253| 55.6             | 1110.2| 0         |
| 7   | ALFDPRAEV | 1BOG     | -48,750| 1.242| 55.8             | 1047.1| 3         |
| 8   | FLLRDQIKL | 1HHK     | -53,450| 1.240| 55.6             | 1099.6| 2         |
| 9   | FLLQDQIKL | 1HHK     | -53,474| 1.240| 55.6             | 1100.4| 2         |
| 10  | ALFDRTAEV | 1BOG     | -48,739| 1.229| 55.6             | 1056.4| 4         |

a PDB entry of the template structure used in predicting the listed sequence.

b Maximum sequence identity with an HLA-A0201-binding peptide in the MHCPEP data base.

c Difference in the solvent-accessible surface area of the unbound versus bound peptide, whose sequence is given in column 2. The solvent-accessible surface area was computed using software Access (Lee and Richards, 1971).

d Number of hydrogen bonds made between the bound peptide side chains and the MHC protein. H-bonds were computed using software WebLab Viewer Lite (Accelrys).

**TABLE III**

Sequence profiles of HLA-A0201 binding peptides in MHCPEP and those predicted by DESIGNER

| Amino acid type | Maximum identity (%) | ΔASA (kcal/mol) |
|-----------------|----------------------|----------------|
| Ala             | 10.3                  | 4.9            |
| Arg             | 5.9                   | 5.9            |
| Asn             | 3.9                   | 6.4            |
| Asp             | 2.1                   | 6.6            |
| Cys             | 2.1                   | 8.1            |
| Gin             | 3.3                   | 3.9            |
| Glu             | 2.8                   | 5.2            |
| Gly             | 5.1                   | 7.0            |
| His             | 2.8                   | 3.9            |
| Ile             | 4.9                   | 2.9            |
| Leu             | 7.4                   | 5.3            |
| Lys             | 8.2                   | 5.0            |
| Met             | 5.1                   | 9.0            |
| Phe             | 4.0                   | 5.1            |
| Pro             | 4.6                   | 6.4            |
| Ser             | 8.7                   | 5.4            |
| Thr             | 3.1                   | 5.8            |
| Trp             | 4.2                   | 5.8            |
| Tyr             | 4.9                   | 5.3            |
| Val             | 4.6                   | 8.4            |

### Footnotes

* Probabilistic entropy, defined as $S(p) = -\sum p \log p$. The larger the $S(p)$ value, the higher the amino acid diversity of the position. The maximum value of $S(p)$ is 2.996 (−log 1/20). The left most column lists the amino acid type. The position-specific amino acid frequencies for the natural peptides of the MHCPEP data base (32) (left) and for the designed peptide sequences (right) were computed as described under “Experimental Procedures,” and are given in percent. All the designed sequences whose energy is within 2 kcal/mol of the minimum energy sequence were used in computing the frequencies.

Fig. 2. These conformations are shown superimposed onto those of the x-ray structures of the corresponding bound peptide in each complex. We see that although the amino acid sequences of the predicted peptides are quite different from those in the two x-ray structures, the side chain conformations of equivalent residues are quite similar and feature very similar values of the χ1 angles. This similarity is particularly striking in positions, such as the anchor residues, for which the predicted and native amino acids are of the same or of closely similar types.

This result cannot be entirely attributed to the fact that the predictions of the peptide amino acid sequences and conformations were performed in the framework of the fixed template, with the latter including the atomic coordinates of native peptide backbone and the full three-dimensional structure of MHC molecule. Indeed the design calculations were not carried out on a single template but on six different templates, in which equivalent protein side chains in the peptide-binding groove exhibit r.m.s. deviations of about 1 Å, as illustrated in Fig. 1. It is well known that the backbone and side chain conformations of an amino acid residue are closely coupled (33, 37, 38), and the similarity of the predicted and native amino acid side chain conformations is thus most likely due to constraints imposed by native peptide backbone templates.

However, the constrained native MHC environment seems to have played a more important role in the selection of the amino acid type at each position. Support for this assumption comes from test calculations in which DESIGNER was applied to the same MHC-peptide templates, still keeping the backbone coordinates (of both protein and bound peptide) fixed, but allowing the MHC side chains to adjust their conformations (but not their amino acid sequence), during the design procedure. This yielded a set of minimum energy peptide sequences, which had higher energies than those listed in Table II and featured none...
of the expected amino acid types at the anchor residues. These peptide sequences were therefore not considered for further analysis, whereas the 10 predicted peptide sequences of Table II were synthesized, and their binding to MHC and ability to impair CD8+ T cell recognition were analyzed.

**Binding of Peptides to HLA-A2 Molecules**—To test binding to HLA-A2, the 10 selected peptides, numbered 1–10, were added at various concentrations to purified HLA-A2 molecules as described under “Materials and Methods.” Two CD8+ T cell epitopes from the influenza virus were also tested. The NP 383–391 peptide (peptide 11) from the viral nucleoprotein which is presented by HLA-B27 was used as a negative control, and the M.58-66 peptide (peptide 12) from the matrix, presented by HLA-A2, was used as a positive control for binding.

The results presented in Fig. 3, A and B, show that at high concentration of $10^{-7}$ and $10^{-6}$ M, all 10 peptides bind significantly to HLA-A2 albeit with different efficiencies. At the low concentration of $10^{-8}$ M, only a subset of the peptides shows significant binding. By comparing several experiments, it was concluded that the efficient binders, given in the order of decreasing relative binding, are peptides 6, 9, 8, 5, 7, and 2, displaying respectively, 131, 116, 111, 98, 83, and 50% of M.58-66 binding used as an internal reference (peptide 12). Interestingly, peptide 6, exhibiting the highest HLA-A2 binding at $10^{-8}$ M, displayed reduced binding at higher concentrations of $10^{-6}$ and $10^{-7}$ M. A similar behavior was observed with the natural M.58-66 cognate peptide (Fig. 3A).

**Stability of HLA-A2-Peptide Complexes**—Because it has been argued that in viral systems immunogenicity correlates better with stability of HLA-peptides complexes (7), we also evaluated the stability at 37 °C of the formed complexes. This assay was performed using the most potent HLA-A2 binders, namely peptides 2 and 5–9. To study their ability to stabilize the HLA-A2 molecule, they were added at a concentration of $10^{-6}$ M to purified HLA-A2 molecules, and maximal binding for each peptide (considered as 100% for calculation) was obtained at time 1 (defined as 0 h). Complexes were then incubated at 37 °C for various times (1, 3, 5, and 24 h), and the amount of remaining complexes was quantified. The maximal HLA-A2 stability was obtained with peptide M.58-66 (peptide 12), then, in decreasing order, with peptides 6, 2, 5, 9, 8, and 7 (Fig. 4). It is noticeable that HLA-A2 complexes formed with peptide 6 were stable over a 24-h period (about 75% remaining versus 90% in the case of the particularly stable HLA-A2-M.58-66 peptide complexes). The NP 383-391 peptide, known not to bind HLA-A2, used here as negative control.

**Inhibition of CD8+ T Cell Responses**—The most potent binders 2 and 5–9 were also assessed as inhibitors of HLA-A2-restricted CD8+ T cell responses. Inhibition of T cell response was visualized in an ELISPOT test detecting IFN-γ secretion. Two types of T cell effectors both restricted to HLA-A2, were used, one recognizing the melanoma-specific MART 26–35 peptide, and the other recognizing the influenza peptide M.58-66. In the test designed to measure the ability of A2 binders to compete with the binding of MART-(26–35) peptide to HLA-A2,
peptides were formed between peptides added at 10^{-6} m and aliquots of 1 µg of purified HLA-A2 molecules. The maximal numbers of complexes (100% binding) were obtained after a stabilization phase at 4 °C (time 1 = 0 h). Results obtained after various incubation times at 37 °C (from 2–5 corresponding to 1, 3, 5, and to 24 h, respectively) are expressed as the percentage of the maximum number of complexes for each peptide.

Interestingly, peptides 2 and 5–9 were tested and ranked according to their decreasing inhibition efficiency. Peptides 9 and 8 were the most potent inhibitors followed by peptides 5, 6, and 2 in this order (Fig. 5A). In the M.58-66 peptide system, among the four peptides tested (peptides 5, 6, 7, and 9), peptides 5 and 9 gave the most significant inhibition, whereas peptides 6 and 7 were marginally efficient (Fig. 5B). As control, we verified that each peptide added alone was not spontaneously recognized by the two types of T cell effectors (data not shown).

Interestingly, peptides 2 and 5 share very similar sequences, the only difference being at position 1, which is Phe in peptide 2 and Tyr in peptide 5. The presence of Phe at position 1 thus seems important for peptide binding to HLA-A2 and for increasing its capacity to inhibit T cell recognition. Peptides 6, 8, and 9, which were found to be very potent binders at low peptide concentration, have also very similar sequences with only one variation at position 4 or 5. It has to be noted that although peptide 6 gave very stable HLA-A2 complexes, it did not display the highest efficiency for inhibiting T cell recognition. In summary, peptides 5, 8, and 9 are the best candidates for inhibition of CD8\(^+\) T responses.

**DISCUSSION**

In this paper we used an automatic procedure to compute the amino acid sequences of peptides that are likely to bind the HLA-A2 MHC class I allele. A first remarkable result described here is that out of the 10 highest scoring peptides selected by our procedure, all were shown to actually bind the expected MHC protein. The six strongest HLA-A2 binders also promoted the assembly of stable HLA-A2-peptide complexes, albeit to varying degrees, and three peptides displayed significant capacities to inhibit CD8\(^+\) T cell recognition (for a summary of these results see Table S1 of the Supplemental Material).

Of the six most active predicted peptides, peptides 2 and 5 were the best ranking candidates for MHC HLA-A2 binding, selected from among a very large number of sequences solely on the basis of the fitness function used by DESIGNER. It is therefore quite satisfying that peptide 5, in particular, exhibits almost the same binding as a natural CD8\(^+\) epitope, the M.58-66 peptide from the influenza virus matrix, and significant inhibitory properties. Indeed, this suggests that our fitness function, which is based on the well established CHARMM22 force field (24) and represents a quantity akin to the peptide binding free energy, is an effective selection criterion for this design problem.

Three other predicted peptides namely, peptides 6, 8, and 9, display similar or somewhat higher MHC binding than the natural epitope M.58-66, and two of these (8 and 9) also exhibit efficient inhibition of T cell recognition. These peptides belong to the second group of peptides in Table II. They represent the low energy sequences selected by DESIGNER, which also display the highest similarity score against the known set of HLA-A0201-binding peptides in the MHCEP data base.

Interestingly, all five lowest energy sequences have been computed from the same template (that of the 1HHK PDB entry), and the values of their DESIGNER free energies differ little (Table II). The most salient difference between the sequences of peptides 6, 8, 9, and those of peptides 1–5, selected on the basis of the DESIGNER free energy alone, is that the former have Leu as the anchor residue at P9, like in the M.58-66 epitope, whereas the latter have Val in this position. Interestingly, this position is completely buried in the peptide-binding groove of MHC (39). Another notable difference is in the amino acid residue at P8. Peptides 6, 8, 9 have a Lys in this position, whereas peptides 1–5 feature a Phe, which has very different chemical properties. These two differences are probably at the origin of the different activity patterns of these peptides.

**Linking Observed MHC-binding Properties to Structural Features**—In an attempt to link observed properties such as...
MHC binding and stabilization to structural features of the predicted peptides, we computed for each predicted peptide sequence and structure the number of intermolecular H-bonds formed by its side chains and the surface area buried upon complex formation. These quantities are listed in Table II.

We could readily establish that the correlation of the activity properties with the number of H-bonds was poor, whereas that with the buried surface area was better. We see indeed that peptide 6, which displays the highest binding and stability activities of all 10 predicted peptides, forms no H-bonds with the MHC molecule, whereas peptide 10, whose biological activity is significantly weaker, forms 4 hydrogen bonds. On the other hand peptide 6 buries the largest surface area of all peptides in the complex, whereas peptide 10 buries the next to lowest surface area. The other two peptides with highest buried surface area are 8 and 9, the next most active peptides following peptide 6. This seemingly good correspondence between the measured activity and the buried surface area is not too surprising in view of the fact that the peptide/MHC interactions are primarily hydrophobic in nature. Interestingly, a near anti-correlation between the buried surface area in the MHC-peptide interface and the number of H-bonds formed between the protein and the peptide side chains is also observed in the set of native MHC-peptide crystal structures used here as templates for the sequence design (Table I).

Fig. 6 displays the conformation of one of the most active predicted peptides (peptide 6) superimposed onto the backbone of the natural M.58-66 epitope bound to the HLA-A2 molecule, as in the 1HHK PDB entry. As expected, the backbone conformation of peptide 6, which also originates from the 1HHI structure, used as template, is quite similar to that of the bound M.58-66 peptide. The N- and C-terminal residues of both peptides are particularly well superimposed, and the side chains of the two anchor residues at P2 and P9 are completely buried in pockets inside the MHC molecule. On the other hand the central portion of the peptide backbones in the vicinity of P5 (a Trp in peptide 6) displays slightly different conformations. In addition, in peptide 6 the Trp side chain at P5 sticks into the solvent and makes no interaction with the MHC molecules. On the other hand, the corresponding Phe side chain in M.58-66 points into the peptide-binding groove and makes hydrophobic interactions with the protein. Despite this difference the M.58-66 epitope buries overall a similar surface area in the native complex as peptide 6 does (1087.3 Å² versus 1112 Å², respectively). Thus in this predicted peptide the N and C termini contribute significantly to HLA-A2 binding, whereas in the M.58-66 epitope, the middle portion contributes significantly.

Origins of the Observed Inhibition of HLA-A2-restricted CD8+ T Cell Response—An interesting property of some of the designed peptides is that they impair activation of the human CTL, mediated by the two types of T cell effectors tested here, those recognizing the MART-(26–35) and the M.58-66 epitopes, respectively, in the context of HLA-A2. Because CTL activation requires recognition of the MHC-peptide complex by the TCR,
it seemed worthwhile to examine the only known crystal structure of an HLA-A2-peptide-TCR ternary complex (see Ref. 5; PDB code 1BD2), for possible clues concerning the features leading to the inhibitory properties of the designed peptides.

In this complex, almost the entire peptide is buried in the TCR/MHC interface, with a majority of the contacts with the peptide being made by the CDR1α and CDR1β loops of the TCR (5). This means that the TCR molecule buries all or nearly all of the parts of the peptide that are not already buried in the binary MHC-peptide complex. These parts consist primarily of the side chains of residues 5 and 8 and to a lesser extent of residues 4 and 6.

In the HTLV-1 Tax peptide in the known structure of ternary complex and in the M.58-66 and MART-(26–35) peptides used here in the test with T cell effectors, position 5 is occupied by aromatic (Phe and Tyr) or Gly and Tyr, position 8 by Thr or Tyr, position 4 by Gly, and position 6 by Val, Pro, or Ile. The correspondence, between the MART peptide, which has 10 residues and the 2 other natural peptides was established using structure superpositions of the peptide backbones in the corresponding crystal structures taken from the PDB. This showed that residues 1–4 and 7–10 of the MART peptide were structurally equivalent, respectively, to residues 1–4 and 6–9 of the HTLV-1 Tax and M.58-66 peptides. But position 5 in the latter peptides was roughly overlapping with positions 5 and 6 of the MART peptide.

In contrast, in the three predicted peptides with highest inhibition activity (peptides 5, 8, and 9), the corresponding positions feature polar and often charged side chains (Tables II or III). Indeed position 5 is Asp in all three peptides, and position 8 is Phe in peptide 5 and Lys in peptides 8 and 9, whereas residues 4 and 6 are, respectively, Gln and Arg in peptide 5, Arg and Gln in peptide 8, and Gln in peptide 9. The selection of such polar and charged side chains in these positions by DESIGNER is energetically advantageous, because the design calculations were performed in the absence of the bound TCR molecules, leaving the side chains at these residues accessible to the solvent.

Thus, the peptides with largest inhibitory activity seem to be those that have a similar or somewhat higher affinity for the HLA-A2 than the natural peptides, and which feature either 2 or 3 charged side chains in the 4 positions mentioned above, that participate in TCR binding.

By using the atomic coordinates of the ternary HLA-A2-peptide-TCR complex, we modeled the conformations of the M.58-66 and MART-26–35 peptides that were used in the inhibition tests and those of the three predicted peptides with highest inhibitory activity of the CTL response. Fig. 7A shows the conformations of the side chains of residues 5 and 8 of the Tax HTLV-1 peptide and the two natural peptides in the environments of ternary complex. The conformations of the corresponding residues in the three designed inhibitory peptides are shown in Fig. 7, B–D, side by side the side chain conformations in the natural and designed peptides, respectively, for residues 4 and 6. Inspection of these figures suggests that the charged and often longer amino acids of the designed peptides are likely to interfere with TCR recognition, due primarily to unfavorable electrostatic and solvation effects. No attempt was made here to optimize the ternary complexes modeled with the designed peptides.

In conclusion our study demonstrates that automatic protein design procedures can be successful in the design of peptides that bind with high enough affinity to one protein so as to impair specific protein-protein recognition mediated by these peptides and involving a second protein. This opens up the prospect for the use of these procedures as a general tool for investigating and modulating protein-peptide and protein-protein interactions.

Although this success may be attributed, at least in part, to advances in the protein design procedures, the constraints imposed by the MHC-binding groove on the choice of peptide side chain conformations and amino acid types have clearly been helpful. These constraints result from the use of accurate structural templates (the high resolution MHC-peptide complexes) and from the fact that the peptides form extensive interactions with the template. We have seen indeed that relaxing these constraints even partially, for instance by allowing the side chains of the MHC-peptide binding groove to adjust their conformation during the design procedure, yielded designed peptides that lacked the anchor residues and, most likely, the expected binding properties as well.

Generalizing the approach described here to cases where the atomic coordinates of the MHC molecule are less accurate, as in the case of models derived by homology (40), or to MHC class II molecules, where the interactions with the peptide are believed to be less constraining, is therefore likely to be more difficult2 and will require further developments.

Acknowledgments—We are grateful to Lorenz Wernisch for help with the initial calculations and to Jean Richelle as well as all the members of the European project for many useful discussions.

REFERENCES

1. Engelhard, V. H. (1994) Curr. Opin. Immunol. 6, 13–23
2. Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L. & Wiley, D. C. (1987) Nature 329, 506–512
3. Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L. & Wiley, D. C. (1987) Nature 329, 512–518
4. Madden, D. R. (1995) Annu. Rev. Immunol. 13, 587–622
5. Garboczi, D. N., Ghosh, P., Utz, U., Fan, Q. R., Biddison, W. E. & Wiley, D. C. (1996) Nature 384, 134–141
6. Sette, A., Vitale, A., Beherman, B., Fowler, P., Nayerseina, R., Kast, W. M., Melief, C. J., Oseroff, C., Yuan, L. & Ruppert, J. (1994) J. Immunol. 153, 5586–5592
7. van der Burg, S. H., Visseren, M. J., Brandt, R. M., Kast, W. M. & Melief, C. J. (1996) J. Immunol. 156, 3308–3314
8. Bush, D. H. & Pamer, E. G. (1989) J. Immunol. 160, 4441–4448
9. Rashinowitz, J. D., Beseen, C., Wurfing, C., Tate, K., Allen, P. M., Davis, M. M. & McConnell, H. M. (1996) Immunity 8, 223–229
10. Reis e Sousa, C., Levine, E. H. & Germain, R. N. (1996) J. Exp. Med. 174, 149–157
11. Singh, R. R. (2000) Curr. Opin. Rheumatol. 12, 399–406
12. Kesh, G. J. & Allen, P. M. (1996) Nature 380, 495–498
13. England, R. D., Kullberg, M. C., Cornette, J. L. & Berzofsky, J. A. (1995) J. Immunol. 155, 4385–4306
14. Fairchild, P. J. (1997) Eur. J. Immunogenet. 24, 155–167
15. Meziere, C., Viguier, M., Dumortier, H., Lo-Man, R., Leclerc, C., Guillet, J. G., Briand, J. P. & Muller, S. (1997) J. Immunol. 159, 3230–3237
16. Oskam, M., Guihard, G., Connan, F., Muller, S., Chaboissier, A., Hoebeko, J., Choppin, J., Briand, J. P. & Guillet, J. G. (1998) J. Immunol. 161, 200–208
17. Weng, Z. & DeLisi, C. (1998) Immunol. Rev. 163, 251–266
18. Rammensee, H., Bachmann, J., Emmerich, N. P., Bachr, O. A. & Stevanovic, S. (1999) Immunogenetics 50, 213–219
19. Gulkota, K., Sidney, J., Sette, A. & DeLisi, C. (1997) J. Mol. Biol. 267, 1258–1267
20. Mamitsuka, H. (1998) Proteins 33, 460–474
21. Roberts, C. G., Meister, G. E., Jasdale, B. M., Lieberman, J., Berzofsky, J. A. & De Groot, A. S. (1996) AIDS Res. Hum. Retroviruses 12, 593–610
22. Wernisch, L., Hery, S. & Wodak, S. J. (2000) J. Mol. Biol. 301, 713–736
23. Jaramillo, A., Wernisch, L., Hery, S. & Wodak, S. J. (2001) Comb. Chem. High Throughput Screen 4, 643–659
24. MacKerrell, A. D., Jr., Bashford, D., Bellott, R. L., Dunbrack, R. L., Jr., Evanseck, J. D., Field, M. J., Fischer, S. Gao, J., Guo, H.-A., Ha, S., Joseph-McCarthy, D., Kuchin, L. K., Kau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodromi, B., Reider, W. E., III, Roux, B., Schnakuire, M., Smith, A. C., Stote, R., Struth, J., Watanabe, M., Wioskiewicz-Kuczera, J., Yin, D. & Karplus, M. (1998) J. Phys. Chem. 102, 3586–3616
25. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. 2

2. Ogata, A., Jaramillo, W., Cohen, J.-P., Briand, F. Connan, J. Choppin, S. Muller, and S. J. Wodak, unpublished results.
& Karplus, M. (1983) J. Comp. Chem. 4, 187–217
26. Ooi, T., Oobatake, M., Nemethy, G. & Scheraga, H. A. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 3086–3090
27. Kuhlman, B. & Baker, D. (2000) Proc. Natl. Acad. Sci. U. S. A. 97, 10383–10388
28. Raha, K., Wollacott, A. M., Italia, M. J. & Desjarlais, J. R. (2000) Protein Sci. 9, 1106–1119
29. Desjarlais, J. R. & Handel, T. M. (1999) J. Mol. Biol. 290, 305–318
30. Gordon, D. B., Marshall, S. A. & Mayo, S. L. (1999) Curr. Opin. Struct. Biol. 9, 599–513
31. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000) Nucleic Acids Res. 28, 235–242
32. Brusic, V., Rudy, G. & Harrison, L. C. (1998) Nucleic Acids Res. 26, 368–371
33. Dunbrack, R. L., Jr. & Karplus, M. (1993) J. Mol. Biol. 230, 543–574
34. Neimark, J. & Briand, J. P. (1993) Pept. Res. 6, 219–228
35. King, D. S., Fields, C. G. & Fields, G. B. (1990) Int. J. Pept. Protein Res. 36, 255–266
36. Choppin, J., Cohen, W., Bianco, A., Briand, J. P., Connan, F., Dalod, M. & Guillett, J. G. (2001) J. Immunol. 166, 6164–6169
37. Tanimura, R., Kidena, A. & Nakamura, H. (1994) Protein Sci. 3, 2358–2365
38. Eisenmenger, F., Argos, P. & Abagyan, R. (1993) J. Mol. Biol. 231, 849–860
39. Madden, D. R., Garboczi, D. N. & Wiley, D. C. (1993) Cell 75, 693–708
40. Sanchez, R. & Sali, A. (1997) Curr. Opin. Struct. Biol. 7, 206–214
41. Lee, B. K. & Richards, F. M. (1971) J. Mol. Biol. 55, 379–400
