Amino Acid Sequence of Cyanogen Bromide Fragment CN-C (Residues 24–98) of the Mouse Histocompatibility Antigen H-2D^d: A Comparison of the Amino-Terminal 100 Residues of H-2D^d, D^b, K^d, and K^b Reveals Discrete Areas of Diversity†

Roderick Nairn,‡ Stanley G. Nathenson,* and John E. Coligan

ABSTRACT: The amino acid sequence of the cyanogen bromide (CNBr) fragment extending from position 24 to position 98 of the H-2D^d murine histocompatibility antigen has been determined by using radiochemical microsequencing techniques. This 75-residue fragment (CN-C) is one of two glycopeptides generated by CNBr cleavage of the extracellular portion of the H-2D^d molecule (H-2D$^d_{papain}$). Determination of this sequence completes the amino-terminal 100 residues of the H-2D^d molecule. The primary structure of CN-C was established by thrombin digestion of isolated CN-C and sequence determination of the three resulting peptides. The COOH-terminal Met and its adjacent residue were determined by sequence analysis of a tryptic peptide which overlaps CNBr fragments C and b4 (residues 99–138). Alignment of the thrombic peptides was accomplished by NH_2-terminal sequence analyses of CN-C and peptides generated by *Staphylococcus aureus* V8 protease digestion of CN-C. The sequence data presented here, together with that already given for H-2D^d [Nairn, R., Nathenson, S. G., & Coligan, J. E. (1980) *Eur. J. Immunol. 10*, 495–503], allow a comparison of the NH_2-terminal 100 residues of the D^d, D^b, K^d, and K^b molecules. Discrete areas of diversity, in particular, one between residues 62 and 83, are obvious. Comparison over some 180 residues of the D^d and K^b molecules reveals a particularly close similarity between these products of a K and a D gene from widely disparate mouse strains.

The major histocompatibility complex (MHC)[1] is a tightly linked cluster of genetically defined regions specifying molecules that regulate immunological reactions [reviewed in Klein (1975, 1979), Shreffler & David (1975), Snell et al. (1976), and Göetze (1977)]. The class I regions (Klein, 1977) in the murine MHC (H-2) encode the class I molecules, i.e., the H-2K, D, and L histocompatibility antigens. A class I molecule exists in the cell surface membrane as a glycosylated protein of M_r ~45 000 associated noncovalently with a nonglycosylated polypeptide chain of M_r ~12 000, β_2-microglobulin [reviewed in Snell et al. (1976)]. These ubiquitous cell surface glycoproteins are the major targets for graft rejection in vivo and for cell-mediated lympholysis in vitro. They also play a major role in the "restriction" of immune cytolytic processes; e.g., killing of virally infected cells requires recognition of the viral antigen in association with specific class I molecules [reviewed in Klein (1979) and Zinkernagel (1979)]. Hence, it is of considerable importance to completely define the structure of these molecules.

The overall molecular features, i.e., chain length and number and position of carbohydrate side chains and disulfide loops, are similar, if not precisely identical, for products of the H-2K, D, or L genes (Ewenstein et al., 1978; Kimball et al., 1980; Nairn et al., 1980a; Coligan et al., 1980). Microsequence analyses of radiolabeled class I molecules [reviewed in Nathenson et al. (1981)] have been utilized successfully to obtain data on the primary structure of a number of H-2K, D, and L molecules. When the data available for H-2K, D, and L molecules are compared with that for the homologous class I molecules of man, the most striking features are the high degree of overall primary structure homology and the clusters of amino acid sequence diversity among long stretches of highly homologous sequence (Orr et al., 1979b; Coligan et al., 1980; Nairn et al., 1980a; Uehara et al., 1980a; Martinko et al., 1980; Maloy et al., 1981; Kimball et al., 1981). Such comparisons of the primary structure data have suggested that the largest cluster of sequence variability exists between residues 61 and 83.

In this report, we document the primary structure of CNBr fragment C (residues 24–98) from the H-2D^d molecule, thereby allowing comparisons of sequence data for residues 1–100 of a pair of alleles from both the K and D regions as well as comparisons between the products of the class I regions of both mouse and man.

Experimental Procedures

Cells, Radiolabeling, and Antisera. The C14 cell line, induced by Abelson virus in a BALB/c mouse, was utilized as the source of radiolabeled H-2D^d antigen (Nairn et al., 1980a). Whereas in our previous studies on H-2 molecules Asp residues have been assigned indirectly [reviewed by Nathenson et al. (1981)], a modification of our labeling methods has permitted us to incorporate enough of this radiolabled amino acid to make direct assignments for Asp residues (F. T. Gates III, E. S. Kimball, and J. E. Coligan, unpublished experiments). The alloantiserum used to detect the H-2D^d specificity H-2.4 was prepared in the combination (*H-2^b* × *H-2^k*)F$_1$ anti *H-2^a*.

Isolation of the H-2D$^d_{papain}$ Molecule. Details of these procedures were given previously (Nairn et al., 1980a). The H-2D^d antigen was prepared by immunoprecipitation from the

† From the Department of Microbiology and Immunology and the Department of Cell Biology, Albert Einstein College of Medicine, Bronx, New York 10461 (R.N. and S.G.N.), and the Laboratory of Immunogenetics, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20205 (J.E.C.). *Received February 12, 1981.* These studies were supported in part by Grants AI-07289 and AI-10702 from the National Institutes of Health (S.G.N.) and by a National Multiple Sclerosis Fellowship, FG-490-A-1 (R.N.).

‡ Present address: Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI 48109.

[1] Abbreviations used: β_2m, β_2-microglobulin; MHC, major histocompatibility complex; Gdn·HCl, guanidine hydrochloride; HPLC, high-performance liquid chromatography; F$_3$CCOOH, trifluoroacetic acid; V8, *Staphylococcus aureus* V8 protease; PTH, phenylthiohydantoin.

0006-2960/81/0420-4739$01.25/0

Table I: Radioactivity Recovered in the Different Preparations of CN-C Thrombic Peptides

radioactive amino acids	initial radioact. in CN-C (cpm)	radioactivity θ1 cpm	θ1 % [a]	θ1 % expect [b]	θ2 cpm	θ2 %	θ2 % expect	θ3 cpm	θ3 %	θ3 % expect	recovery of radioact. (%) [c]
LYR	76000	11950	33	25	13400	37	45	11300	31	30	48
K	30000	<100		0	9600	100	100	<100		0	32
E(P) [d]	10480	<100		0	2400	52	56	2200	48	44	44
HITVWF	158200	21800	34	21	26300	42	37	15200	24	42	40
A	48800	7700	32	40	9800	40	40	6700	28	20	50
P	91200	<100		0	4300	37	33	7300	63	67	13
S [e]	24500	9300	56	50	3500	21	25	3800	23	25	68
G(S) [f]	12700	3590	42	41	2620	31	38	2270	27	21	67
Q(E) [g]	13100	3500	55	44	2900	45	51	<100		5	49
N	43800	8800	30	25	7900	27	25	12800	43	50	67
D	12500	<100		0	1180	27	25	3150	73	75	35

[a] Percent = (radioactivity in peptide/total radioactivity recovered in θ1, θ2, and θ3) × 100. [b] % radioactivity was calculated from the composition determined by the sequence (correction for relative specific activity made in the case of multilabel preparations). [c] (Sum of cpm in all thrombic peptides/initial cpm) × 100. [d] Approximately 40% of the radioactivity in the Glu preparation was Pro. [e] No radioactive Gly was found in the Ser preparation. [f] Approximately 15% of the radioactivity in the Gly preparation was Ser. [g] Less than 10% of the radioactivity in the Gln preparation was Glu. In this case no radioactive Pro was detected.

glycoprotein pool, i.e., that fraction of the NP-40 lysate of the cells which binds to columns of the lectin from *Lens culinaris* and is then eluted by 0.1 M α-methyl mannoside. Treatment of the immunoprecipitates with papain allowed the subsequent isolation of the H-2D$^d_{papain}$ molecule (M_r ~37000) by chromatography of the products of papain digestion on columns of Sephadex G-75 (2.5 × 100 cm) equilibrated in 1 M HCOOH.

Preparation of CNBr Fragment C (CN-C). The H-2D^d molecule was digested with CNBr as described previously for H-2K^b (Ewenstein et al., 1978), and the products of digestion were separated by chromatography on a column of Sephacryl S-200 (2 × 190 cm) in the presence of 6 M Gdn·HCl (Nairn et al., 1980a). Desalting of peptides was accomplished on columns of Sephadex G-15 (2.5 × 40 cm) equilibrated with 2 M HCOOH.

Thrombin Digestion. Thrombin (B grade, lot no. 703053 and 903415, Calbiochem, La Jolla, CA) was further purified from either of the separate lots used as described by Lundblad (1971). The esterase activity of the preparation against *p*-tosyl-L-arginine methyl ester (TAME), as determined by the spectrophotometric method of Hummel (1959), showed a ΔA_{247}/(min·mL) equal to 4.4 and 9.4, respectively, at 23 °C. Digestions were carried out as described by Uehara et al. (1980a). In brief, CN-C in the presence of 1–5 mg of carrier horse heart cytochrome *c* was dissolved in 1.5 mL of 0.1 M NH_4HCO_3 and incubated at 37 °C for 20 h with 0.15 mL of thrombin solution. The reaction was terminated by the addition of 0.2 mL of 88% formic acid, and the products of digestion were separated by chromatography on a column (0.9 × 175 cm) of Sephadex G-50 (Superfine grade) in the presence of 1 M HCOOH. Similar results were obtained with either thrombin preparation.

Preparation of Tryptic Overlap Peptides. [^{35}S]M/[^{3}H]-HITVWF-labeled H-2D^d molecules were isolated by immunoprecipitation and separated from β_2m by chromatography on a column of Sephadex G-75 (2.5 ×100 cm) in the presence of 1 M HCOOH, in this case only, without prior digestion with papain. After reduction and carboxyamidomethylation, H-2D^d was digested with TPCK-trypsin (Worthington Biochemicals, Freehold, NJ) as described by Brown et al. (1974), desalted by chromatography on Sephadex G-15 in the presence of 2 M HCOOH, and lyophilized. Following solubilization in 2 M HCOOH, this material was chromatographed on a column (1.5 × 190 cm) of Sephadex G-25 (Superfine grade) in 2.0 M HCOOH. Fractions were combined to make pools on the basis of the presence of ^{35}S radioactivity and, after lyophilization, were dissolved in 0.5 mL of 0.05% F_3CCOOH and fractionated by HPLC. An E-M Lichrosorb (Merck, Darmstadt, West Germany) C-8 reverse-phase column (4.6 × 250 mm) was empolyed at a flow rate of 1 mL/min on a Waters Model 240 HPLC (Waters Associates, Inc., Milford, MA). Gradient elution from 0.05% F_3CCOOH in H_2O to 0.05% F_3CCOOH/40% *n*-propyl alcohol over a period of 80 min was used to develop the chromatogram after an initial isocratic elution for 5 min.

Preparation of V8 Peptides. Staphylococcal aureus V8 protease (lot no. 0877, Miles Laboratories, Inc., Elkhart, IL) was used without further treatment. CNBr fragment C (containing 1.0 mg of carrier cytochrome *c*) was dissolved in 1.0 mL of 0.5% NH_4HCO_3, pH 8.0, and 0.02 mL of V8 solution (1.0 mg/mL in 0.5% NH_4HCO_3, pH 8.0) was added. The solution was incubated at 37 °C for 24 h and then the reaction stopped by addition of 0.12 mL of 88% formic acid. The products of digestion were separated by chromatography on a column (0.9 × 220 cm) of Sephadex G-50 (Superfine grade) equilibrated in 2.0 M HCOOH.

Automated Amino Acid Sequence Analysis. Extensive details of this methodology have been given previously (Coligan et al., 1978, 1979; Uehara et al., 1980a). Briefly, a Beckman 890-C automated sequencer was utilized for Edman degradations with Beckman Programs 102974, 121078, and 21980 in the presence of several milligrams of horse heart cytochrome *c*. The latter two programs were used after a cold trap was added to the sequencer. Polybrene (Aldrich Chemical Co., Milwaukee, WI) was added to the sequencer cup for peptides less than 50 residues in length to prevent premature washout (Klapper et al., 1978). Radiolabeled PTH amino acids were identified by cochromatography with unlabeled standard PTH amino acid derivatives by HPLC, using the isocratic elution conditions described by Gates et al. (1979).

Results

Isolation of Thrombic Peptides from CN-C of H-2D^d. In order to carry out the sequence studies reported here, the H-2D^d molecule, radiolabeled with the combination of amino acids documented in Table I, was isolated. After CNBr digestion, fragment C (extending from residue 24 to residue 98) was purified by chromatography on a Sephacryl-S200 column in 6 M Gdn·HCl. CN-C, a glycopeptide having an apparent

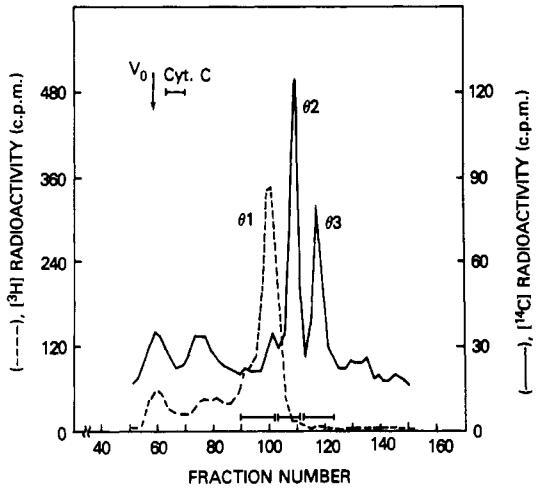

FIGURE 1: Chromatography of a thrombic digest of the [^{14}C]Arg- and [^{3}H]fucose-labeled CN-C fragment of H-2D^d on a column of Sephadex G-50 (Superfine grade). The column size was 0.9 × 175 cm, and the solvent was 1.0 M HCOOH. The fraction size was 0.75 mL and the flow rate 3 mL/h; 150-μL aliquots were taken for determination of radioactivity by liquid scintillation counting. Fractions were combined as indicated by the horizontal bars to obtain peptides θ1, θ2, and θ3. The migration position of the marker protein cytochrome *c* (M_r 12 400) is indicated. V_0 indicates the void volume of the column.

molecular weight slightly less than that of cytochrome *c*, was previously demonstrated to be a pure peptide by NH_2-terminal sequence analyses (Nairn et al., 1980a).

As shown in Figure 1, three major fragments (θ1, θ2, and θ3) were generated from the glycopeptide CN-C after reaction with thrombin, a proteolytic enzyme of the blood clotting system (Seegers, 1967). Fragment θ1 was the only fragment which incorporated radioactive fucose (Figure 1), indicating that it contained the carbohydrate side chain known to be present in this CNBr fragment. Some aggregated material at the void volume as well as some undigested material which eluted prior to the cytochrome *c* marker was generally present, but the major peaks of radioactivity had highly reproducible elution positions in the different separations (Figure 1).

The recovery of radioactivity in each of the thrombic peptides from the 11 separate preparations is given in Table I. In general, there was excellent agreement between the relative radioactivity recovered in each of the thrombic peptides and that expected from their composition as inferred subsequently from the sequence (Table I). The range of recovery of starting radioactivity in the three thrombic fragments was 32–68%, with the exception of [^{3}H]Pro-labeled material where only 13% of the starting radioactivity in CN-C was recovered as thrombic peptides. The reason for this one exceptionally low recovery is not known, but there was no large peak of undigested material or contaminating material observed.

Sequence Studies of the Thrombic Peptides. The amino acid sequences of the peptides were determined. The method of identification of each residue (Table I-S) and the radioactivity recovered in each step during these sequence analyses (Figure 1-S) are provided (see paragraph at end of paper regarding supplementary material).

Partial sequence analyses of θ3 (residues 24–48) served to confirm many of the residues determined during our previous NH_2-terminal sequence analyses of CN-C (Nairn et al., 1980a) but also extended the data (N at residues 30 and 42, for example). Complete sequence data were obtained for θ2 (residues 49–79) which like θ3 terminated in an Arg residue, consistent with the known specificity of thrombin. The complete sequence of thrombic peptide θ1 was determined except

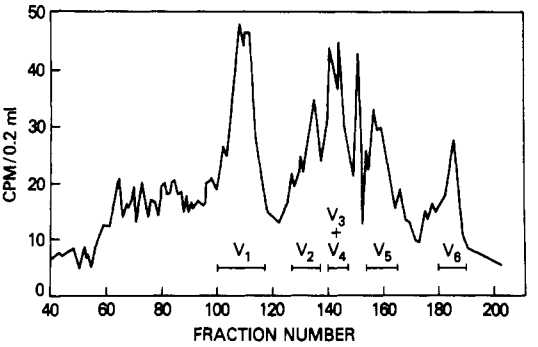

FIGURE 2: Chromatography of a *Staphylococcus aureus* V8 protease digest of the [^{3}H]HITVWF-labeled CN-C fragment of H-2D^d on a column of Sephadex G-50 (Superfine grade). The column size was 0.9 × 220 cm, and the solvent was 2.0 M HCOOH. The fraction size was 2.0 mL and the flow rate 7.2 mL/h; 200-μL aliquots were taken for determination of radioactivity by liquid scintillation counting. Fractions were combined as indicated by the horizontal bars to obtain peptides V1–V6.

for the Trp-Met residues at positions 97–98 which were determined by sequence analysis of a tryptic overlap peptide as discussed later.

An Asn residue was assigned to position 86 of θ1 for a number of reasons. First, although the peptide labels in Asn (Table I), no Asn residues were detected in θ1 and all positions except 86 were assigned. The assignment of Asn-86 is further supported by the fact that θ1 is known to be a glycopeptide (Figure 1) and the Asn-Gln-Ser sequence at residues 86–88 would constitute the recognition sequence (Neuberger & Marshall, 1969) for N-glycoslyation. Glycosylated residues are not extracted by butyl chloride during sequence analysis (Waterfield & Bridgen, 1975), which would explain the failure to detect Asn at this position. In addition, H-2K^b (Uehara et al., 1980a), H-2K^d (Kimball et al., 1981), H-2D^b (Maloy et al., 1981), HLA-B7, and HLA-2 (Orr et al., 1979a,b) are known to have an N-linked carbohydrate side chain at residue 86.

Other Enzymatic Cleavages of CN-C. CN-C labeled in [^{3}H]HITVWF was digested with *Staphylococcus aureus* V8 protease, and the fragments generated were separated on a column of Sephadex G-50 in 2.0 M HCOOH (Figure 2). Peptides V1–V6 were isolated and their sequences determined. These peptides were useful in confirming prevous assignments, but, most importantly, sequence analysis of peptide V1 allowed confirmation of the alignment of peptides θ1 and θ2 (see Figure 5).

Digestion with trypsin of [^{35}S]M/[^{3}H]HITVWF-labeled, -reduced, and -carboxyamidomethylated H-2D^d molecules generated peptides which were separated by chromatography on columns of Sepahdex G-25 in 2.0 M HCOOH (Figure 3). The peptides eluting in peak TO1 were further fractionated by reverse-phase HPLC (Figure 4), and the sequence of the [^{35}S]M-containing peptides was determined. One of these peptides (TO1-1) enabled the positive assignment of the COOH-terminal Met residue of CN-C as well as the adjacent Trp-97 which had not been detected during sequence analysis of θ1 (Figure 5), probably due to premature washout. In addition, Val-103 was assigned from this peptide, confirming that peptide CN-b4 immediately follows CN-C (see Discussion and Figures 5 and 7).

Discussion

Utilization of the proteolytic enzyme thrombin to generate suitably sized peptides (~30 residues) for automated sequence analysis allowed determination of the entire sequence of CN-C

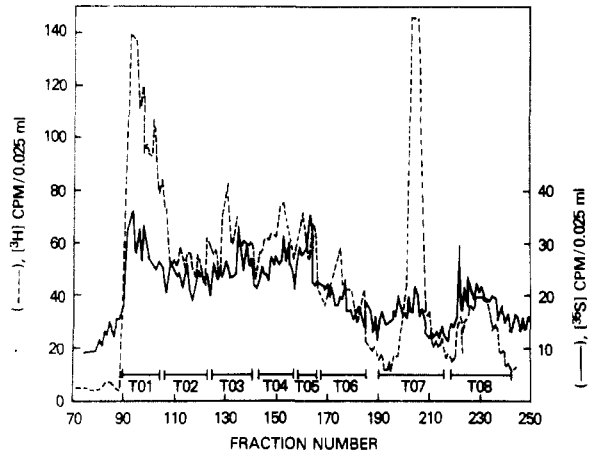

FIGURE 3: Chromatography of a trypsin digest of [^{3}H]HITVWF/[^{35}S]M-labeled H-2D^d on a column of Sephadex G-25 (Superfine grade). The column size was 1.5 × 190 cm, and the solvent was 2.0 M HCOOH. The fraction size was 1.3 mL and the flow rate 6.0 mL/h; 25-μL aliquots were taken for determination of radioactivity by liquid scintillation counting. Fractions were combined as indicated by the horizontal bars to obtain the tryptic peptides TO1-TO8.

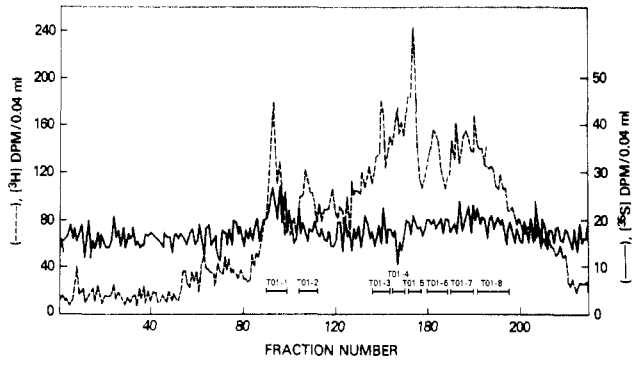

FIGURE 4: Reverse-phase high pressure liquid chromatogram of a [^{3}H]HITVWF/[^{35}S]M-labeled tryptic peptide (T01) on an E-M Lichrosorb C-8 column. The column size was 4.6 × 250 mm, and it was developed after an initial isocratic elution for 5 min with a gradient run over a period of 80 min from 0.05% F_3CCOOH in H_2O to 0.05% F_3CCOOH/40% *n*-propyl alcohol. The fraction size was 0.4 mL, and the flow rate on a Waters Model 240 HPLC was 1.0 mL/min at 25 °C; 40-μL aliquots were taken for determination of radioactivity by liquid scintillation counting. Fractions were combined as indicated by the horizontal bars to obtain the tryptic peptides TO1-1–TO1-8.

(Figure 5). The alignment of thrombic peptides θ3 and θ2 was accomplished by NH_2-terminal sequence analyses of CN-C. Such analyses previously established the NH_2-terminal residue of CN-C to be position 24 in the intact H-2D^d molecule (Nairn et al., 1980a). Alignment of thrombic peptides θ2 and θ1 was from sequence analysis of the V8 peptide CN-C-V_1 (Figure 5). The placement of cyanogen bromide fragment b4 (residues 99–138) at the COOH terminal of CN-C is confirmed here by sequence analysis of the tryptic overlap peptide TO1-1 (Figure 5).

The radioactivity recovered in the thrombic peptides appears to accurately reflect the relative compositions of these peptides calculated from the known sequence (Table I). For example, θ2 is rich in Glu, Lys is found only in θ2, Gln is absent from θ3, and Pro and Glu are absent from θ1. Similarly, the expected percentages of Ser radioactivity calculated from the amino acid composition after the sequence was known are 50:25:25 in θ1:θ2:θ3. The observed percentages were 56:21:23. After correction for specific activity differences, multilabel preparations also show a satisfactory parallel between the ratios expected and those observed (Table I), which in the absence of classical composition data gives reassurance as to the reliability of radiochemical methods.

A comparison of the NH_2-terminal 100 residues is made for four murine and two human class I molecules in Figure 6. A striking aspect of this comparison is the high degree of sequence homology between the H-2 and HLA molecules (~70%). When the sequences of any of the H-2 molecules are compared, slightly higher homology (75–87%) is observed in all cases. As indicated in the bottom two lines of Figure 6, residues at 65 positions are common to all four H-2 molecules and 56 residues (or 86%) are common to all six H-2 and HLA molecules. In addition, all six class I molecules examined so far have an N-linked carbohydrate side chain attached at residue 86. These findings strongly suggest that the class I molecules of these two species evolved from a common ancestral gene.

For the NH_2-terminal 100 residues, the homology of D^d to D^b, D^d to K^d, and D^d to K^b is 81%, 75%, and 87%, respectively, and the homology of D^d to either B7 or A2 is 67%. The largest number of sequence differences is found between residues 61 and 83 when either the murine or human molecules are compared. When D^d is compared to D^b, to K^d, and to K^b over the stretch of sequence between residues 61 and 83, the observed homologies are 65% , 57%, and 83%, respectively. This area of greater than average diversity has been noted previously in the comparison of H-2 and HLA molecules (Orr et al., 1979b; Coligan et al., 1980; Nairn et al., 1980a; Uehara et al., 1980a; Martinko et al., 1980; Maloy et al., 1981; Kimball et al., 1981).

Another prominent, although less extensive, segment of differences among the four murine molecules occurs between residues 95 and 100. Noticeably, position 97 is the only position in the NH_2-terminal 100 residues of the H-2 molecules where four different amino acids have been assigned. Similar diversity is not found for residues 95–100 when the two HLA molecules are compared to each other.

The sequence comparisons lead to certain generalizations. For example, no "K-ness" or "D-ness" is readily apparent in the NH_2-terminal sequences for 100 residues of the two K and the two D gene products. Close inspection of these sequences reveals only two positions which might be "locus specific", i.e., residues 30 and 89 (Figure 6). Also, D^d is more like K^b than it is like D^b. In fact, K^b and D^d are the most homolgous (87%) of the four murine molecules examined.

There are no data to suggest that H-2K products are more like HLA-B products than HLA-A products, or vice versa (Figure 6). Hence, notions of close evolutionary relationships between particular loci in different species are untenable at this time.

The total sequence information known for the H-2D^d glycoprotein is given in Figure 7. For the 179 positions determined, the sequence of H-2D^d is 88% homologous with that of the recently completed sequence of the papain fragment of H-2K^b (Uehara et al., 1980b; Martinko et al., 1980). Most of the sequence differences detected so far are located in the NH_2-terminal 100 residues (13/21, see Figure 6). The remaining differences are scattered throughout the molecule and occur at residues 114, 116, 141, 144, 145, 152, 163, and 256 (Figure 7).

The difference between the D^d and K^b molecules at position 116 (Tyr in K^b, Phe in D^d) is one which has also frequently been detected in comparisons of the parental K^b molecule with the K^b molecule isolated from strains of mice having mutations in the *H-2K^b* gene. Thus, there is a Tyr→Phe interchange at residue 116 in 5 out of 12 mutants of the *H-2K^b* series [reviewed in Nairn et al. (1980b)]. Close to this highly variable

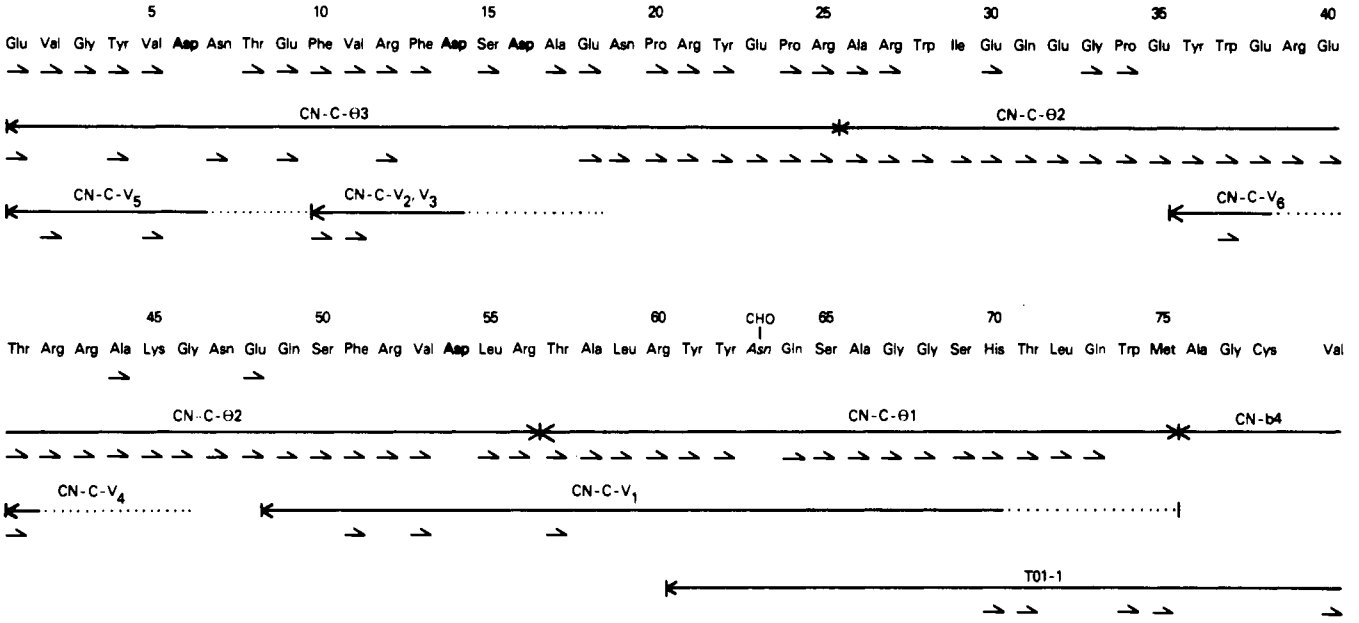

FIGURE 5: Amino acid sequence of CN-C (residues 24–98) of the H-2D^d molecule. The numbers above the sequence refer to the residue positions in fragment CN-C. The first line of arrows indicates those residues indentified by NH_2-terminal sequence analyses of CN-C. The arrows below the solid line depicting the position of the thrombic peptides indicate which amino acids were identified in thrombic peptides (θ), staphylococcal protease V8 peptides (V), and tryptic overlap peptides (TO). The residue (Asn-63) in italics was assigned indirectly. CHO represents the carbohydrate side chain.

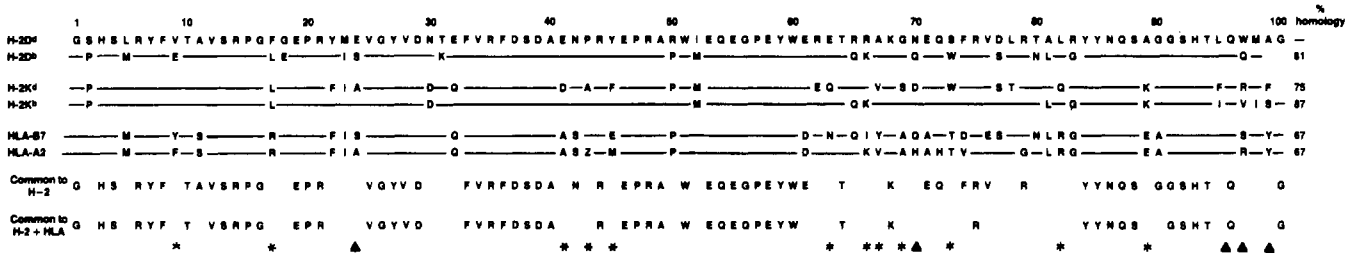

FIGURE 6: Comparison of the NH_2-terminal 100 residues of amino acid sequence of four murine and two human histocompatibility antigens. Residues identical with those in the sequence of H-2D^d are indicated by solid lines; differences are identified. The standard one-letter symbols are used. (▲) indicates a position where more than two different amino acids are assigned among the four H-2 molecules. * indicates a position where more than two different amino acids are assigned among the six molecules. Data for H-2D^b were taken from Maloy et al. (1981), for K^d from Kimball et al. (1981), and for K^b from Uehara et al. (1980). Data for HLA-B7 and HLA-A2 were from Orr et al. (1979a,b).

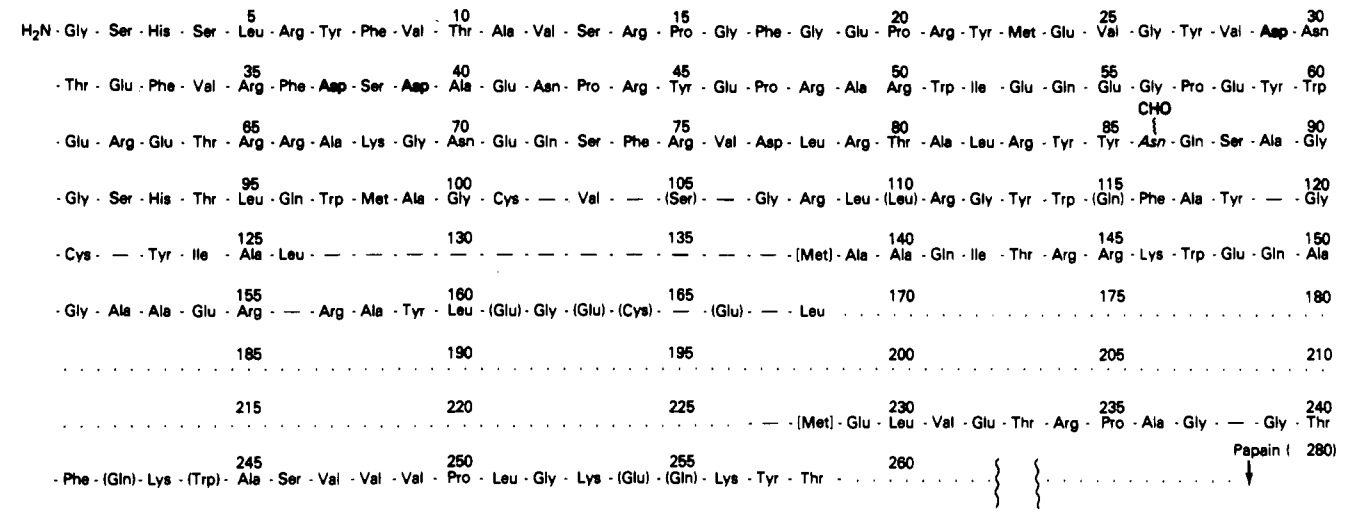

FIGURE 7: Complete sequence information available for the mouse histocompatibility antigen H-2D^d. The standard three-letter symbols for the amino acids are used. Residues in parentheses are either Met in positions assumed from the CNBr cleavage patterns or tentative assignments. The unidentified residues are represented by a single dash.

residue at positions 116, H-2D^d (Figure 7), H-2K^b (Uehara et al., 1980a), and HLA-B7 (Orr et al., 1979a,b) share an amino acid sequence, Gly-Arg-Leu-Leu-Arg-Gly (at residues 107–112), which is a potential palindrome for 18 bases at the nucleic acid level. In support of this, Steinmetz et al. (1981) have shown that in an H-2 cDNA clone (pH-2III) isolated from an *H-2^k* cell line, there is a palindromic base sequence for residues 107–112. The Gly in D^d, K^b, and HLA-B7 at position 107 is apparently replaced with Trp in this case, making the base sequence a palindrome over 16 bases rather

than the potential 18 mentioned above. That a site of frequent mutation should occur closely apposed to such a structural feature in the nucleic acid is intriguing. Perhaps some nucleic acid processing event occurs close to this sequence.

Why D^d and K^b should be particularly closely related (as compared to the other murine class I molecules for which there are substantial sequence data) is difficult to explain. These two molecules are from different gene series and from widely disparate mouse strains. The two stains of origin for these particular antigens are BALB/c and C57BL/6, the former probably originating in North West Europe and the latter probably derived from original stock imported from China (Klein, 1975). Clearly, the sequence data point to the fact that there are no simple relationships among the genes of the class I regions of the MHC. This close homology between K^b and D^d was noted previously (Shimada & Nathenson, 1969) with respect to their overall biochemical properties.

The presumed alleles, K^b and K^d (75% homology) and D^b and D^d (81% homology), have been found to possess significant sequence differences in the first 100 residues (Figure 6), as had been predicted from earlier peptide map comparisons (Brown et al., 1974). While it is tempting to speculate that the large number of sequence differences observed between alleles is directly related to the significant differences detected between them by a variety of serological and cellular test systems, it should be noted that the absolute number of sequence differences may be considerably less important than where the differences are located. Many of the differences could be immunologically silent. For example, the *H-2* mutants examined have only one or a few amino acid substitutions in their H-2 molecules, yet they can give very strong reactions in immunological test systems [reviewed in Nairn et al. (1980b)]. As yet, however, we have insufficient data to allow us to speculate reliably about potentially immunodominant areas of the class I molecule.

In summary, it is clear that the sequence data obtained so far raise fundamental questions concerning the evolution of the class I regions of the MHC. It is difficult to explain the observed structural relationships by a "simple" mechanism of primordial gene duplication and the subsequent accumulation of differences by mutation after speciation occurred. Rather, the sequence data suggest that the observed polymorphism has been generated by more complex mechanisms. The accumulation of amino acid sequence data, such as that presented here, together with data from nucleic acid sequence analyses of the MHC genes will hopefully lead to an understanding of the structure and evolution of the MHC.

Acknowledgments

It is a pleasure to acknowledge the secretarial assistance of Catherine Whelan and Morag Nairn. Mark Flocco, Diane McGovern, Selma Schumacher, Pat Spinella, Joanne Trojnacki, and Robert Valas provided expert technical assistance. Helpful discussions of the data and the manuscript with our colleagues Ed Kimball, Tom Kindt, Lee Maloy, John Martinko, Larry Pease, Hiroshi Uehara, and Kazushige Yokoyama were appreciated. We are particularly indebted to Hiroshi Uehara for advice with the thrombin experiments and to Frederick Gates for help with the Asp labeling.

Supplementary Material Available

A table outlining the method(s) of identification for each of the residues in the peptides prepared from CN-C and a figure plotting the recovery of radioactivity above background (as $\log_2$ cpm) in each step of the automated amino acid sequence analyses of the peptides comprising CN-C (3 pages). Ordering information is given on any current masthead page.

References

Brown, J. L., Kato, K., Silver, J., & Nathenson, S. G. (1974) *Biochemistry 13*, 3174–3178.

Coligan, J. E., Kindt, T. J., Ewenstein, B. M., Uehara, H., Nisizawa, T., & Nathenson, S. G. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 3390–3394.

Coligan, J. E., Kindt, T. J., Ewenstein, B. M., Uehara, H., Martinko, J. M., & Nathenson, S. G. (1979) *Mol. Immunol. 16*, 3–8.

Coligan, J. E., Kindt, T. J., Nairn, R., Nathenson, S. G., Sachs, D. H., & Hansen, T. H. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 1134–1138.

Ewenstein, B. M., Nisizawa, T., Uehara, H., Nathenson, S. G., Coligan, J. E., & Kindt, T. J. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 2909–2913.

Gates, F. T., III, Coligan, J. E., & Kindt, T. J. (1979) *Biochemistry 18*, 2267–2272.

Göetze, D., Ed. (1977) in *The Major Histocompatibility System in Man and Animals*, Springer-Verlag, New York.

Hummel, B. C. W. (1959) *Can. J. Biochem. Physiol. 37*, 1393–1399.

IUPAC-IUB Commission on Biochemical Nomenclature (1968) *J. Biol. Chem. 243*, 3557.

Kimball, E. S., Maloy, W. L., Martinko, J. M., Nathenson, S. G., & Coligan, J. E. (1980) *Mol. Immunol. 17*, 1283–1291.

Kimball, E. S., Nathenson, S. G., & Coligan, J. E. (1981) *Biochemistry 20*, 3301–3308.

Klapper, D., Wilde, C. W., & Capra, J. D. (1978) *Anal. Biochem. 85*, 126–131.

Klein, J. (1975) in *Biology of the Mouse Histocompatibility-2 Complex*, Springer-Verlag, New York.

Klein, J. (1977) in *The Major Histocompatibility System in Man and Animals* (Göetze, D., Ed.) pp 339–378, Springer-Verlag, New York.

Klein, J. (1979) *Science (Washington, D.C.) 203*, 516–521.

Lundblad, R. L. (1971) *Biochemistry 10*, 2501–2506.

Maloy, W. L., Nathenson, S. G., & Coligan, J. E. (1981) *J. Biol. Chem. 256*, 2863–2872.

Martinko, J. M., Uehara, H., Ewenstein, B. M., Kindt, T. J., Coligan, J. E., & Nathenson, S. G. (1980) *Biochemistry 19*, 6188–6193.

Nairn, R., Nathenson, S. G., & Coligan, J. E. (1980a) *Eur. J. Immunol. 10*, 495–503.

Nairn, R., Yamaga, K., & Nathenson, S. G. (1980b) *Annu. Rev. Genet. 14*, 241–277.

Nathenson, S. G., Uehara, H., Ewenstein, B. M., Kindt, T. J., & Coligan, J. E. (1981) *Annu. Rev. Biochem.* (in press).

Neuberger, A., & Marshall, R. D. (1969) in *Symposium on Foods—Carbohydrates and their Roles* (Schultze, H. W., Cain, R. F., & Wrotstad, R. W., Eds.) p 115, Avi Publishing Co., Westport, CT.

Orr, H. T., Lopez de Castro, J. A., Lancet, D., & Strominger, J. L. (1979a) *Biochemistry 18*, 5711–5720.

Orr, H. T., Lopez de Castro, J. A., Parham, P., Ploegh, H. L., & Strominger, J. L. (1979b) *Proc. Natl. Acad. Sci. U.S.A. 76*, 4395–4399.

Seegers, W. H. (1967) in *Blood Clotting Enzymology*, Academic Press, New York.

Shimada, A., & Nathenson, S. G. (1969) *Biochemistry 8*, 4048–4062.

Shreffler, D. C., & David, C. S. (1975) *Adv. Immunol. 20*, 125–195.

Snell, G. D., Dausset, J., & Nathenson, S. G. (1976) in

Histocompatibility, Academic Press, New York.
Steinmetz, M., Frelinger, J. G., Fisher, D., Hunkapiller, T., Pereira, D., Weissman, S. M., Uehara, H., Nathenson, S. G. & Hood, L. (1981) *Cell (Cambridge, Mass.) 24*, 125–134.
Uehara, H., Ewenstein, B. M., Martinko, J. M., Nathenson, S. G., Coligan, J. E., & Kindt, T. J. (1980a) *Biochemistry 19*, 306–315.
Uehara, H., Ewenstein, B. M., Martinko, J. M., Nathenson, S. G., Kindt, T. J., & Coligan, J. E. (1980b) *Biochemistry 19*, 6182–6188.
Waterfield, M. D., & Bridgen, J. (1975) in *Instrumentation in Amino Acid Sequence Analysis* (Perham, R. N., Ed.) p 41–71, Academic Press, New York.
Zinkernagel, R. M. (1979) *Annu. Rev. Microbiol. 33*, 201–213.

Mechanism of Ribonucleic Acid Chain Initiation. Molecular Pulse-Labeling Study of Ribonucleic Acid Synthesis on T7 Deoxyribonucleic Acid Template†

Nobuo Shimamoto,‡ Felicia Y.-H. Wu, and Cheng-Wen Wu*

ABSTRACT: The mechanism of the productive initiation of RNA synthesis in vitro by *Escherichia coli* RNA polymerase holoenzyme was investigated by using DNA from the T7 deletion mutant ΔD111 as a template under conditions such that RNA chains are initiated exclusively from the A1 promoter. Kinetic studies by a non-steady-state method revealed that the binding of the first two nucleoside triphosphates, corresponding to the 5′-terminal and penultimate nucleotides of the RNA transcript, to the enzyme–promoter complex during the productive initiation is rapid and ordered, namely, ATP binds first followed by UTP. The same mechanism was also observed for productive initiation at the D promoter. The time course of the incorporation of the first four 5′-terminal nucleotides into the RNA synthesized from the A1 promoter was studied by a fast kinetic technique. Under our experimental conditions, the times for the half-maximal incorporation of ATP, UMP, CMP, and GMP, each at an initial concentration of 0.1 mM, were 0.03 (or less), 0.16, 0.35, and 0.40 s, respectively. These incorporations were completed within 1 s and maintained for up to 90 s at a constant level which was dependent on the initial nucleotide concentration. The half-saturation concentrations for ATP, UTP, CTP, and GTP were 80, 88, 20, and 7 μM, respectively. At saturating nucleotide concentrations, the level of incorporation was approximately equal to the promoter concentration. We have found that productive initiation at the A1 promoter is activated by the third nucleotide, CTP, and the fourth nucleotide, GTP, the former being a more potent activator than the latter. The observation that mono- and diphosphates of these two nucleotides are also activators of the productive initiation suggests that the activation is not due to the phosphodiester bond formation. Kinetic analysis indicates that the activator binds to a regulatory site on the enzyme–promoter complex and exerts its action at a step after the formation of the first phosphodiester bond. Furthermore, GTP inhibits the abortive initiation at the A1 promoter. A minimal mechanism of RNA chain initiation consistent with all our results is proposed. RNA polymerase binds to a promoter site to form an open-promoter complex. The binding of the first two nucleotides to this binary complex is rapid and ordered. A phosphodiester bond is then formed to yield a dinucleotide, with the release of pyrophosphate. In the presence of a regulatory nucleotide, corresponding to the third or fourth nucleotide, the translocation of the enzyme along the DNA template is facilitated to form a productive initiation complex which is ready for elongation of an RNA chain. In the absence of the regulatory nucleotide, the dinucleotide is released from the enzyme–DNA complex, and the initiation is aborted. Discrimination between the productive and abortive initiation pathways by a regulatory nucleotide may play an important role in the control of specific RNA synthesis and in the enhancement of the fidelity of transcriptional initiation.

Gene transcription by bacterial RNA polymerase can be divided into several steps (Goldthwait et al., 1970; Chamberlin, 1976): (a) binding of enzyme to a promoter on the DNA template, (b) initiation of RNA synthesis, (c) elongation of RNA chains, and (d) termination of RNA synthesis. Although a wealth of information has been accumulated in the past decade about prokaryotic gene transcription, we are still far from understanding the molecular mechanisms by which RNA polymerase carries out this complicated series of reactions.

The initiation of RNA synthesis involves the binding of both the 5′-terminal and the penultimate nucleoside triphosphates to the RNA polymerase–promoter complex as well as the formation of the first phosphodiesterase bond (Goldthwait et al., 1970). Earlier studies on initiation have indicated that the 5′-terminal nucleoside triphosphate is primarily a purine nucleotide (Maitra & Hurwitz, 1967) and that purine nucleotides are required at relatively higher concentrations for initiation than for elongation (Anthony et al., 1969). Thus initiation can be distinguished from elongation by the purine nucleotide requirement and by the relatively high apparent K_m derived from steady-state kinetic analysis (Goldthwait et

† From the Department of Biochemistry, Albert Einstein College of Medicine, Bronx, New York 10461, and the Department of Pharmacological Sciences, State University of New York at Stony Brook, Long Island, New York 11794. *Received January 29, 1981.* This work was supported in part by Research Grants GM 28069 from the National Institutes of Health and NP 309G from the American Cancer Society.
* Address correspondence to this author at the Department of Pharmacological Sciences, State University of New York at Stony Brook.
‡ Present address: Faculty of Integrated Arts and Sciences, Hiroshima University, Hiroshima, Japan 730.

0006-2960/81/0420-4745$01.25/0