

4071

# **Molecular Modeling**

## **From Virtual Tools to Real Problems**

**Thomas F. Kumosinski, EDITOR**

*U.S. Department of Agriculture*

**Michael N. Liebman, EDITOR**

*Amoco Technology Company*

Developed from a symposium sponsored  
by the Division of Agricultural and Food Chemistry  
at the 205th National Meeting  
of the American Chemical Society,  
Denver, Colorado,  
March 28–April 2, 1993

# Molecular Modeling of Apolipoprotein A-I Using Template Derived from Crystal Structure of Apolipophorin III

Eleanor M. Brown, Thomas F. Kumosinski, and Harold M. Farrell, Jr.

Eastern Regional Research Center, Agricultural Research Service,  
U.S. Department of Agriculture, 600 East Mermaid Lane,  
Philadelphia, PA 19118

Plasma lipoprotein particles are the major vehicles for lipid transport in the circulatory systems of animals. Apolipoprotein A-I (apo A-I), the primary high density lipoprotein (HDL), serves as a cofactor in the esterification of cholesterol and mediates the transport of cholesterol esters to the liver for utilization. Structure-function studies of the apolipoproteins are hindered by the limited crystallographic data available. Apolipophorin III (apo Lp-III) of flying insects is responsible for delivery of lipids for utilization by flight muscles. The 2.5-Å resolution structure for apo Lp-III from the African migratory locust, *Locusta migratoria*, is used in the development of a template for the construction of partial models of canine, human and avian apo A-I. Residues 7 to 156 of apo Lp-III were aligned with residues 72 to 236 of apo A-I using alanine as a spacer residue. Four of the five helices of apo Lp-III were preserved in the apo A-I model. Helix 4, the longest of the original helices and the one with the most inserted residues, was separated into two shorter helices connected by a nonhelical strand. Amphipathic character was similar in all models. Electrostatic interactions were more important in the apo A-I models, resulting in increased stability of about 6 kcal/mol/residue.

Lipids serve as structural components of membranes and as energy depots in the cells of all living organisms. Thus, the transport of these sparingly water soluble molecules through the aqueous environment of an organism is of major biochemical importance. Discrete complexes of lipids and proteins, the lipoproteins of mammals and birds or their companion complexes, lipophorins of insects, have evolved as primary transport vehicles for lipids. Lipoprotein particles consist of a core of neutral lipids, stabilized by a surface monolayer of polar lipids complexed with one or more proteins. The protein components of lipoproteins, called apolipoproteins, are amphipathic in nature, having both hydrophobic and hydrophilic regions. Although the number of discrete apolipoproteins and their specific functions vary, comparable proteins are often found in different species. Apolipoproteins A-I (apo A-I) and apo B are respectively, the major protein components of high density (HDL) and very low density (VLDL) lipoproteins of mammals and birds. Most mammalian lipoprotein classes also contain

apolipoprotein E (apo E), a protein that functions with apo B in directing the delivery and redistribution of lipids to cells that express receptors (1,2).

Research over the past 30 years has resulted in complete amino acid sequences and a wealth of detail concerning the biosynthesis, physical characteristics and metabolism of apolipoproteins. Acquisition of structural data at the molecular level has, however, been hindered by the noncrystalline character of most apolipoproteins. This lack of crystallographic data inspired the development of predictive techniques for approximating suitable structures. Correlation of global secondary structures determined spectrophotometrically with local secondary structures predicted from amino acid sequences have been used to suggest structural motifs. Spectrophotometric studies, mainly circular dichroism, show high levels of helical structure for apolipoproteins in aqueous solution and still higher levels when lipids are included in the system (3-5). Through analysis of apolipoprotein primary structure, Boguski and coworkers (6) established a pattern of repeating sequences punctuated by highly conserved proline residues for this family of highly homologous proteins (6-10). Algorithms for the prediction of secondary structure in globular proteins (11,12) and for the analysis of hydrophathy patterns (13,14) were applied to the development of models for the apolipoproteins (15,16). A high potential for the formation of amphipathic helices, characterized by opposing hydrophobic and hydrophilic faces, has been established as a common motif (6,17-19).

Mammalian apo A-I is synthesized in the liver and the intestine (20). Among its functions are the activation of lecithin:cholesterol acyltransferase (LCAT) and the transport of cholesterol from peripheral tissues to the liver for metabolism (21). Synthesis of avian apo A-I occurs in peripheral tissues as well as in the liver, particularly at the time of hatching (22). In addition to LCAT activation it is thought to function in the mobilization of yolk lipids (22). The biochemistry and physiology of insects resemble in a general fashion the corresponding metabolic pathways of vertebrates (23). Two exceptions are that the insect has an open circulatory system in which the hemolymph is enclosed by membranes and a specialized tissue, the fat body, that combines many of the functions of the vertebrate liver and adipose tissue. The majority of hemolymph lipids are found in a single lipoprotein particle, called lipophorin, that is synthesized in the fat body and circulates in the hemolymph. All lipophorins have been observed to have at least two apolipophorins (apo Lp-I and apo Lp-II). A third apolipophorin (apo Lp-III) that functions in the transport of lipid through the hemolymph from storage depots to flight muscles during prolonged flight is found mainly in insects that use lipids to fuel flight (23).

Crystal structures at 2.5-Å resolution have now been published for the 18-kDa apo Lp-III from the African migratory locust, *Locusta migratoria* (24) and the 22-kDa N-terminal, receptor-binding domain of human apo E (25). Despite the functional differences of apo E and LpIII, the crystal structures are remarkably similar. Nolte and Atkinson (26) used the data from these crystal structures in an integrated approach that included statistical analyses, information theory and sequence homology to predict secondary structure and possible tertiary folding patterns for apo A-I and apo E-3 in lipid environments. Delivery of lipids from storage depots to the site of metabolism, the common functional characteristic of apo Lp-III and apo A-I, would argue for a degree of structural similarity. In this study, computer-based molecular modeling has been applied to the search for common structural elements in these functionally related proteins.

## Methods

**Sequence Comparison.** Amino acid sequences for apo Lp-III, human apo A-I (HuA-I), canine apo A-I (DgA-I) and chicken apo A-I (ChA-I) are available in the Protein Identification Resource (PIR) (27). The overall similarity of apo Lp-III to apo

A-I and the amphipathic helical potential of both was established previously (28). The positions of residues 7 to 156 of apo Lp-III were resolved in the crystal structure of this protein (24). IALIGN, an interactive alignment program distributed with the PIR (27) was used to align residues (7 - 156) of apo Lp-III with each of the apo A-I proteins. The best alignment used one to four residue gaps to spread this 150-residue segment of apo Lp-III over a 165-residue segment of apo A-I.

Segments of apo Lp-III that were identified as helical from the crystal structure (24) were evaluated with the "strip of helix" template of Vazquez and coworkers (29) to determine amphipathic potential. The best fit for each helix was then applied to the corresponding segment of DgA-I, HuA-I and ChA-I using the convention of Brasseur (30) that classifies Pro as unique, Gly as equivalent to any other residue, Asn, Asp, Arg, Gln, Glu, His, Lys, Ser, Thr as hydrophilic residues and Ala, Ile, Leu, Met, Phe, Trp, Tyr, Val as hydrophobic residues.

**Molecular Models.** A three dimensional representation of apo Lp-III (residues 7 - 156) derived from a residue by residue reading of the positions of  $\alpha$ -carbons in the published crystal structure (24) was constructed using molecular modeling software (SYBYL, v. 5.5) (31) on a SGI 4D/35 workstation (Silicon Graphics, Mountain View, CA). (Reference to a brand or firm name does not constitute endorsement by the U.S. Department of Agriculture over others of a similar nature not mentioned.) This model was refined by SYBYL-directed energy minimization using the Kollman United Atom force field (32,33). As a template for apo A-I, the model called apo Lp-IIIa was derived from the apo Lp-III structure by the insertion of alanine residues at each of the gap positions identified by IALIGN.

Because the DgA-I sequence most closely matched that of apo Lp-IIIa this model was constructed first. The SYBYL command "mutate residue" was used to substitute residues 72 - 236 of DgA-I into the apo Lp-IIIa sequence. The DgA-I model was then subjected to energy minimization by the SIMPLEX method (31), a rapid nonderivative calculation that moves one atom at a time to reduce any overlapping of atoms that may be present in the initial structure. Refinement by the SIMPLEX method was continued until no atom in the model had a calculated force on it greater than 1000 kcal/mol  $\text{\AA}$ . The conjugate gradient method was then used with the Kollman United Atom force field (32,33) to further refine the model. A distance dependent dielectric function was used in the force field to implicitly account for the dielectric screening due to solvent molecules (which are not explicitly included in the models). When the total energy for the model was reduced to -1000 kcal/mole, the partially refined DgA-I model was used as a template for mutation to the HuA-I and ChA-I models.

Refinement of models was by SYBYL-directed energy minimization carried out *in vacuo* using a root-mean square (rms) force of 0.01 kcal/mole  $\text{\AA}$  as the cutoff value for ending the minimization. At the completion of this minimization step, models were subjected to a 300K molecular dynamics simulation for 1psec followed by an additional minimization step using the same limits as the initial minimization.

Conformational details of the energy minimized models were obtained by analysis of the dihedral angles  $\phi$  and  $\psi$ . An  $\alpha$ -helical segment was identified when both  $\phi$  and  $\psi$  were between  $-30^\circ$  and  $-75^\circ$  for six or more consecutive residues. Single residues with dihedral angles outside these limits could be accommodated as distortions in longer helical segments. This definition of  $\alpha$ -helical conformation was applied to the refined conformation of apo Lp-III (24) so that in this study the amino acids of apo Lp-III in helices 1 through 5 (H1 - H5) are residues 7 - 25 (H1), residues 35 - 59 (H2), residues 70 - 91 (H3), residues 95 - 121 (H4) and residues 136 - 155 (H5). Schiffer-Edmunson (34) amphipathic helical wheel projections (not shown) were used in the evaluation of the amphipathic character of the refined helical segments. The angle between residues along an  $\alpha$ -helical backbone was assumed to be  $100^\circ$  (34). Circular plots from an end-on perspective of the helix were constructed, the magnitude

in degrees of each hydrophobic sector was measured, and the average hydrophobicity of this hydrophobic face was calculated using the Eisenberg normalized hydropathy scale (13).

## Results

**Sequence Comparisons.** Amino acid residues 7 - 156 of apo Lp-III are contained in five long  $\alpha$ -helices connected by short loops (24). Figure 1 shows the alignment of apo Lp-IIIa with DgA-I, HuA-I and ChA-I. The best alignment produced by the program IALIGN for apo Lp-III with DgA-I, HuA-I and ChA-I used one to four-residue gaps to align residues 7 - 156 of apo Lp-III with residues 72 - 236 of DgA-I or ChA-I or residues 73 - 237 of HuA-I. A Pro-Pro repeat near the N-terminus of HuA-I causes this sequence to be shifted by one position relative to DgA-I or ChA-I (26). For simplicity, residues in the modeled apo A-I fragments will be referred to as though all were numbered 72 - 236. Although gaps are an accepted feature of sequence alignments, the model builder must find another way to space out the residues. Alanine was chosen as the spacer residue in building apo Lp-IIIa because it has a small, nonionic side chain lacking notable functionality except for a high probability of being found in helical structures. At 16 positions, the residues in Lp-IIIa, DgA-I, HuA-I and ChA-I are identical, 17 additional identities between DgA-I and apo Lp-IIIa as well as numerous functional identities are observed.

The amphipathic potentials of sequence fragments proposed for helical conformation may be found in Table I. Values of 1 or 0 were assigned for residues that did or did not fit, respectively, the "strip of helix" template (29). Average values for the 5 helices of apo Lp-III were optimized by varying the starting position for the "strip of helix" formalism (29). This optimum analysis was then applied to each of the other sequences. The highest hydropathic potentials are calculated for sequences of residues having hydrophobic side chains at alternating third or fourth positions. The five helices of apo Lp-III with 19, 25, 22, 27, and 20 residues each are notably longer than those commonly found in globular proteins, and thus provide a useful test of the "strip of helix" template. The amphipathic potentials for the five apo Lp-III helices fall in a narrow range between 0.67 and 0.80. The insertion of alanine residues in an arbitrary fashion to fill the gap positions lengthened helices 3, 4, and 5 to 24, 33 and 21 residues, respectively, still within the range of helical lengths (19 to 35 residues) found in apo E (25). The additional alanine residues had relatively little effect on the amphipathic potentials (0.61 to 0.80). The range of amphipathic potentials was greater for the apo A-I sequences (0.45 to 0.88). However, all values for helices 1, 2, 3 and 5 were greater than 0.5, suggesting that each of these helices if stable would be suitable for a lipid-aqueous interface.

**Table I. Amphipathic potentials of predicted helical segments in apolipoprotein models**

Model	H1		H2		H3		H4		H5	
	#res	AP								
Lp-III	19	0.79	25	0.80	22	0.77	27	0.70	21	0.67
Lp-IIIa	19	0.79	25	0.80	24	0.67	33	0.61	22	0.64
DgA-I	19	0.58	25	0.84	24	0.75	33	0.48	22	0.68
HuA-I	19	0.53	25	0.88	24	0.71	33	0.45	22	0.73
ChA-I	19	0.53	25	0.88	24	0.58	33	0.52	22	0.82

Amphipathic potentials (AP) are the best average value for a helical segment of (#res) residues measured with the "strip of helix" (29).



**Molecular Modeling.** Energy minimized models of apo Lp-III, apo Lp-IIIa and the apo A-I proteins are shown in Figure 2a and Color Plate 3. Apo Lp-III and Lp-IIIa are compared in Figure 2 to show the effects of alanine residues on the complete model (2a) and on individual helical segments H3, H4, and H5 (2b, 2c, and 2d). The template model apo Lp-IIIa and the apo A-I protein fragments were compared in terms of stability, amphipathic character and general appearance with the crystallographically determined structure of apo Lp-III (24). All models including apo Lp-III were subjected to the same refinement procedures. Values of the potential energy per residue for each helical segment and for each entire sequence fragment are listed in Table II. The apo Lp-III structure was slightly more stable with an energy of -22 kcal/mol/residue than apo Lp-IIIa (-20 kcal/mol/residue). Stabilization energies of the apo A-I models are essentially identical (-24 kcal/mol/residue) and greater than either apo Lp-IIIa or apo Lp-III. In Color Plate 3, lateral views of the energy minimized Lp-III, DgA-I, HuA-I and ChA-I emphasize the potential for electrostatic interactions, while end-on views emphasize the arrangement of hydrophobic sidechains.

**Table II. Energetic evaluation of the refined models**

	Lp-III	Lp-IIIa	DgA-I	HuA-I	ChA-I
Number of residues	150	165	165	165	165
Energy, kcal					
bond stretching	20.7	23.1	28.9	29.2	31.8
angle bending	123.4	154.0	189.6	208.5	212.2
torsional	198.1	235.7	308.3	324.6	340.9
out of plane	26.1	35.8	37.7	41.5	44.5
1-4 van der Waals	218.7	235.6	247.2	260.2	270.3
van der Waals	-993.0	-1094.3	-1074.8	-1131.8	-1155.6
1-4 electrostatics	1684.0	1813.3	1497.1	1566.7	1524.1
electrostatic	-4497.8	-4567.3	-5214.7	-5182.6	-5208.4
H-bond	-69.3	-70.2	-64.5	-69.8	-76.3
Total Energy, kcal	-3289.0	-3234.3	-4045.2	-3953.6	-4016.4
kcal/mol/residue	-21.9	-19.6	-24.5	-24.0	-24.3
H1 (residues)	(7-25)	—	(72-90)	(73-88)	(72-90)
kcal/mol/residue	-16.1		-18.5	-16.6	-17.4
H2 (residues)	(35-59)	—	(100-124)	(101-125)	(100-124)
kcal/mol/residue	-16.1		-20.1	-19.4	-19.2
H3 (residues)	(70-91)	(70-91)	(135-158)	(136-159)	(135-158)
kcal/mol/residue	-17.6	-20.7	-18.4	-16.2	-17.6
H4 (residues)	(96-121)	(96-121)	(166-196)	(167-197)	(166-196)
kcal/mol/residue	-15.5	-13.8	-16.7	-16.9	-17.6
H5 (residues)	(136-155)	(136-155)	(207-235)	(208-236)	(207-235)
kcal/mol/residue	-15.8	-15.4	-14.0	-13.4	-14.9

These energy calculations, based on the sequence segments initially assigned to H1 - H5 (see Figure 1), are given to illustrate the stability of the structures. Sequences for helices 3 - 5 (H3, H4, H5) of apo Lp-IIIa contain inserted alanine residues, see Methods.

NOTE: The color plates can be found in a color section in the center of this volume.

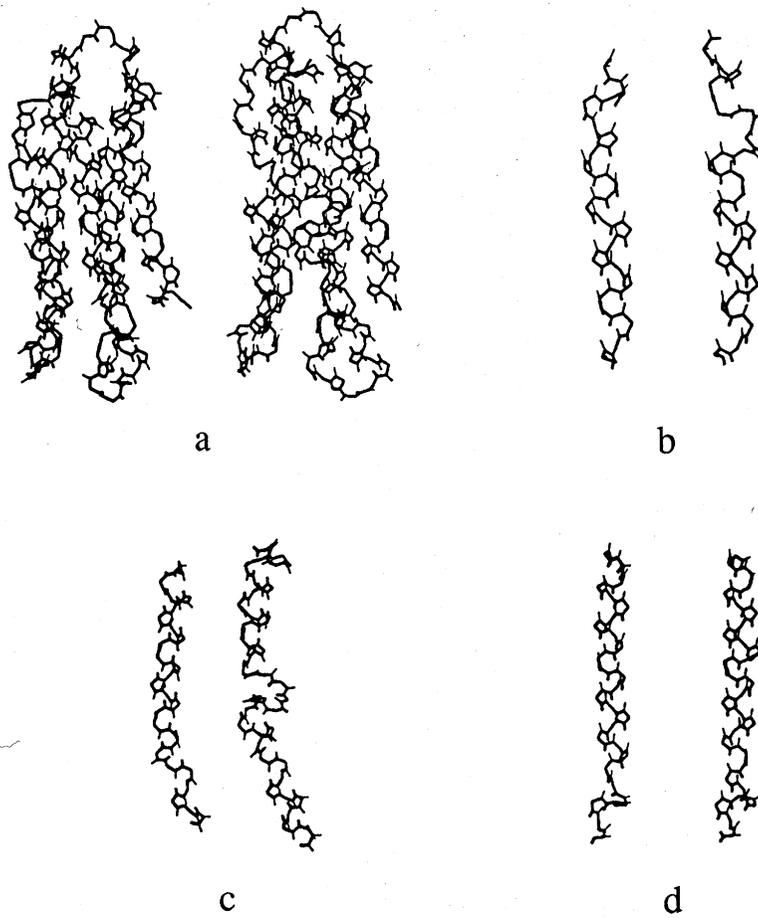


Figure 2. Energy refined models of apo Lp-III and the template apo Lp-IIIa constructed by inserting alanine residues into gap positions identified when the apo Lp-III sequence was aligned with canine, human and chicken apo A-I sequences. Backbone structures for the complete models are in panel (a). The effects of inserted alanine residues on H3, H4 and H5 are displayed in panels (b), (c) and (d) respectively. In each display the apo Lp-III is to the left of the apo Lp-IIIa.

**Helix 1.** With 19 residues, H1 was the shortest helical segment. The substitution of apo A-I sequences into this helix had less effect than might have been expected. The C-terminal residue of H1 was Glu in all of the models. A second Glu found in position four of apo Lp-III, DgA-I and HuA-I was replaced by Leu in ChA-I (Figure 1). Nevertheless, the helical character of H1 (residues 72 - 90) in both DgA-I and ChA-I withstood energy refinement to a stabilizing energy of -17 to -18 kcal/mol/residue (Table II). The amphipathic character of H1 was diminished as the hydrophobic sector was narrowed from 140° to 100° and the average hydrophobicity of this sector was reduced from 1.0 to 0.8 (Table III). Although the corresponding HuA-I sequence differed from DgA-I only in the replacement of Val by Gly at position nine, upon refinement this helix was shortened by three residues from the C-terminal end, the hydrophobic sector reduced to 60°, and the average hydrophobicity to 0.7.

**Table III. Amphipathic analysis of energy refined helices**

Model	Lp-III	Lp-IIIa	DgA-I	HuA-I	ChA-I
H1	N7-E25	N7-E25	D72-E90	D73-K88	E72-E90
Hb sector	140°	140°	100°	60°	100°
Av. Hb	0.98	0.98	0.78	0.70	0.81
H2	P35-S59	P35-S59	L100-E124	L101-E125	L100-E124
Hb sector	100°	100°	100°	100°	100°
Av. Hb	0.80	0.80	0.70	0.89	0.95
H3	S70-T91	S70-S85	Q137-R150	E136-R149	L135-L151
Hb sector	140°	120°	100°	100°	120°
Av. Hb	0.93	0.79	0.95	0.88	0.82
H4					
H4A	A96-S121	Q98-T107	D168-K181	Q172-E179	D167-R181
Hb sector	100°	140°	80°	100°	80°
Av. Hb	0.56	0.91	1.10	0.94	1.18
H4B	—	Q117-a121d	S187-S196	A190-A194	P186-V195
Hb sector	—	60°	100°	—	60°
Av. Hb	—	0.45	0.45	—	0.66
H5	E136-V156	L134-A155	L213-I232	L214-T237	E212-L235
Hb sector	120°	80°	100°	100°	100°
Av. Hb	0.74	0.53	1.05	0.96	0.86

The size of the hydrophobic (Hb) sector of a helix was determined from helical wheel projections (34). The average hydrophobicity (Av. Hb), of this sector was calculated by the method of Eisenberg (13). Single letter designations are used for amino acid residues, (a121d) designates the fourth alanine residue inserted between residues 121 and 122 of Lp-III.

**Helix 2.** The 25-residue H2 was marked by Asp in position two, Glu in position ten and Leu in the first position past the C-terminal residue in all models. The length of this helix and the size of the hydrophobic sector ( $100^\circ$ ) were conserved in all of the apo A-I models. The average hydrophobicity of the hydrophobic sector showed minor variation among the sequences but with values of 0.7 to 0.95, the comparison with a hydrophobicity of 0.8 for apo Lp-IIIa was generally favorable. As in H1, the apo A-I models of H2 were marginally more stable than apo Lp-III (Table II).

**Helix 3.** The first gap position in the sequence alignment required the insertion of two alanine residues between residues sixteen and seventeen of H3 in apo Lp-III. Although alanine residues were inserted into the sequence with  $\alpha$ -helical conformational angles ( $\phi, \psi = -58^\circ, -47^\circ$ ) (31), the result after refinement was a kink in the helix that essentially halted the progression of that helix (Figure 2b). Thus, the sequence contributing to H3 was increased in length from 22 to 24 residues by the addition of two alanine residues, while the helical structure was reduced in length to 16 residues. In the alignment of the 24-residue H3 sequence of apo Lp-IIIa with apo A-I two identities were observed, Leu in position six and His in position twenty. In the refined apo Lp-IIIa model this histidine residue was no longer in the helical structure. The shortened H3, with its associated nonhelical strand, was energetically slightly more stable than H3 of Lp-III (Table II). In refined apo A-I models H3 was further shortened to 14 or 15 residues by the removal of one or two residues from either end. At  $100^\circ - 120^\circ$  the hydrophobic sectors of the helical wheel projections for apo Lp-IIIa and the apo A-I proteins were less than the  $140^\circ$  for apo Lp-III, but still within the range of apo Lp-III helices as were the average hydrophobicity values (0.8 - 0.9) (Table III). H3 in the apo A-I models was energetically more similar to apo Lp-III than to apo Lp-IIIa (Table II).

**Helix 4.** The most dramatic effects on the model structure were in the region of H4. With 27 residues, H4 was the longest of the native helices and coincidentally the only one for which our criteria and that of Breiter and coworkers (24) agreed exactly as to position and length. The alignment of apo Lp-III with the apo A-I sequences dictated the insertion of two alanine residues in the loop region between H3 and H4, six residues into the H4 sequence, and four more to the C-terminus of H4 (Figure 1). The aggregate effect of these insertions on H4 of the apo Lp-IIIa model was to unwind the N-terminal turn of this helix and break H4 into two helical segments, H4A and H4B, separated by a nonhelical strand (Figure 2c). The four alanine residues following the C-terminal serine residue of H4 were incorporated into H4B. The energy of this final conformation in Lp-IIIa was  $-21$  kcal/mol/residue suggesting stability similar to H4 of apo Lp-III (Table II). The sequence alignment (Figure 1) shows several identical residues among the proteins in the region of H4 -- Pro in the N-terminal position, Asp in position three and Leu in positions five, eight and twelve. After energy minimization, the N-terminal Pro and at least two adjacent residues had lost their helical character. Both DgA-I and ChA-I preserved longer helical segments, H4A (14 or 15 residues) and H4B (10 residues), with a shorter nonhelical strand (4 or 5 residues) than did Lp-IIIa. In contrast to the 13-residue nonhelical strand in Lp-IIIa, composed mostly of residues that generally favor helix formation, the five-residue strand in DgA-I contains three Gly residues. In the HuA-I model, the helical strand equivalent to H4A was only eight residues long being shortened from both ends relative to DgA-I and ChA-I. This shortened H4A structure was followed by a ten-residue nonhelical strand, containing two glycine residues, and then by five residues with  $\alpha$ -helical dihedral angles, a sequence too short to evaluate by the helical wheel method. In the apo A-I models H4A had a relatively small ( $80^\circ - 100^\circ$ ) hydrophobic sector with an unusually high average hydrophobicity (0.9 - 1.2). H4B, on the other

hand, was much more hydrophilic with an average hydrophobicity for the hydrophobic sector of 0.5 - 0.6.

**Helix 5.** The 20-residue H5 of apo Lp-III required a single alanine insertion in an already alanine rich region for alignment with the apo A-I proteins. In all refined models, the start of H5 was shifted two or three residues in the N-terminal direction. H5 in the refined models became a 20 to 23-residue helix with a two or three-residue kink just past the site of the alanine insertion in apo Lp-IIIa (Figure 2d). H5 in the DgA-I model was shortened by four residues from the C-terminal end. In other apo A-I models the helical conformation was maintained to within one residue of the C-terminal end.

**Loops.** Loops that separate helical segments play an essential role in determining intrahelical interactions that stabilize the entire structure. The insertion of an additional three residues in the region of the very short loop (3 residues) separating H3 from H4 caused the first turn at the start of H4 to unwind, although it did not change the orientation of H4. Residues inserted at the beginning of the loop separating H4 from H5 in apo Lp-III were incorporated into the H4B of apo Lp-IIIa, again having little effect on the orientation of the following helical segment (Figure 2a). Substitution of apo A-I sequences into the apo Lp-IIIa model had very little effect on the conformation of the loop regions despite the relatively few identical residues in these regions. The average hydrophobicity of amino acid residues in the loop regions of these models were in a narrow range between 0.3 and -0.6, that is neither particularly hydrophilic nor hydrophobic. As may be seen in Color Plate 3, there are many more ionized side chains in apo A-I than in apo Lp-III. The loop separating H2 from H3 contains 4 or 5 charged side chains in each of the proteins. Other loops of apo Lp-III contain one or no ionizable side chains, while those loops in apo A-I contain up to 7 ionizable side chains.

**Role of Proline.** Proline residues tend to have a greater effect on local secondary structure in a molecule than any other single residue. Of six proline residues in the crystalline fragment of apo Lp-III (24) Pro33 and Pro129 are in loops, Pro35 and Pro95 are located at the N-termini of helices while Pro118 and Pro120 are located within H4. The  $\phi$  and  $\psi$  angles ( $-52^\circ$  to  $-60^\circ$  and  $-23^\circ$  to  $-60^\circ$ ) for these six proline residues are more similar to those generally observed for residues in  $\alpha$ -helical structures than to angles typically associated with proline residues. In the apo A-I models Pro98 and Pro208 are aligned with Pro33 and Pro129 of apo Lp-III in loop regions. Pro35 is replaced by a leucine residue in apo A-I with no effect on initiation of H2. The Gln94-Pro95-Ala96-Asp97 sequence that initiates H4 in apo Lp-III is replaced by ala-Pro-Ala-ala-Asp in apo Lp-IIIa and Ala-Pro-Tyr-Ser-Asp in apo A-I (Figure 1) with the introduction of a  $\gamma$ -turn centered on the proline residue similar to the proline based turns in the modeling of milk proteins (35). This turn shifts the start of H4 to at least Asp167 (Figure 1). The helical Pro-X-Pro sequence (residues 118 - 120) of apo Lp-III is not found in apo A-I, however, Pro120 and Pro208 in H2 and H3 respectively of apo A-I have  $\alpha$ -helical  $\phi$ ,  $\psi$  angles ( $-52^\circ$ ,  $-35^\circ$ ).

## Discussion

Pinker and coworkers (36) found that the effects on conformation and conformational stability of alanine substitutions in the helices of myoglobin could not be reliably predicted. In a study using mutant myoglobins, destabilization could be predicted when bulky internal residues were replaced, but results were variable when substitution was in the capping region of a helix. Surface substitutions in mid-helix were less detrimental than at the ends, but because of the disruption of interactions

between neighboring residues, the effects on conformation and stability were not completely predictable. Insertion of alanine residues into a helical segment, had the additional effect of displacing formerly neighboring residues and disrupting the stabilizing interactions of those residues. Insertion of alanine residues with helical conformational angles into a nonhelical loop region might be expected to cause a major disruption of conformation. In fact, the effect in these less structured loop regions was less than in the already helical segments. Had it been feasible to add an entire turn of helical structure at each insertion point in an existing helix, the result in terms of maintaining the helix might have been better. Aside from the effects on individual helical and loop segments detailed above, the stabilization energy (Table II) for this rather arbitrary model differed little from that of the apo Lp-III model.

The extent to which a mutant sequence can maintain the conformation, stability and properties of the template used for its construction gives some clue as to the requirements for a given structure. The signature characteristic of apolipoproteins is a series of helical segments connected by short loops (6,19). Similarities between apo A-I sequences are much greater than those between apo Lp-III and apo A-I, and one would expect excellent results in building a model for ChA-I or DgA-I based on an experimentally determined model of HuA-I. However, no such model is available at this time. An obvious difference (Color Plate 3) between apo Lp-III and apo A-I is in the number of ionizable side chains available for electrostatic interactions. The increased stabilization of apo A-I models relative to apo Lp-III is almost entirely due to improvements in the electrostatic energy (Table II).

In comparative studies of human and chicken apo A-I, circular dichroism spectra show a 10 to 15% greater amount of helical structure for the chicken protein relative to human apo A-I (3-5). In the present study, the loss of H4B in the HuA-I model combined with a greater tendency to unwind the ends of other helices resulted in 10 - 12% fewer residues in helical conformations for HuA-I when compared with DgA-I and ChA-I models.

The use of alanine, a helix stabilizing residue, as the spacer residue was undertaken with the expectation that helical structure would be less affected than if residues from an apo A-I sequence were inserted directly into the Lp-III structure. In fact, the insertion of alanine residues in the helical segments of Lp-III had a disrupting effect on helical structure that, to some extent, was ameliorated when apo A-I sequences were substituted into the Lp-IIIa structure. The insertion of any residue into a stable helical structure shifts the relative positions of other residues in at least two turns of the helix resulting in disruption stabilizing electrostatic and hydrophobic interactions. If the inserted residue can participate in new helix stabilizing interactions the disruption may be minimal. Alanine does not participate in the electrostatic interactions that contribute most to favorable energies for the apo A-I models.

The N-terminal residues 1 - 71 of apo A-I were ignored in these models because that portion of the sequences showed little homology with Lp-III (24) and had little predicted helical structure (26,30,34). The importance of this less helical N-terminal region would appear to be in the activation of LCAT, as several of the epitopes critical for the activation of LCAT are located in the N-terminal third of HuA-I (37,38). As the library of experimentally determined apolipoprotein structures is expanded, other models, better suited for apo A-I, that include the nonhelical regions will undoubtedly become apparent.

Both Nolte and Atkinson (26) and Brasseur and coworkers (30) have recently used molecular modeling techniques to develop models for apo A-I based largely on analysis of the amino acid sequence with respect to repeating segments, hydrophobic potential and predicted secondary structures modified by experimentally estimated conformations. Their models each predict six or seven helical regions beginning with residue 68 of the human apo A-I amino acid sequence. In our models of canine and chicken apo A-I the positions of H1, H2, H4A and H4B are in reasonable agreement

with the locations of helices predicted by others (26,30). H3 in our model is shifted several residues toward the C-terminus of the protein but does overlap the other predictions. Initially much of the region between residues 145 and 162 predicted (26,30) to be weakly helical, was modeled as helical, but the structure did not withstand energy refinement.

Total energy for each of the apo A-I models is -24 kcal/mole/residue, significantly better than the -20 kcal/mole/residue calculated for the energy minimized apo Lp-IIIa model. The much larger number of electrostatic interactions in apo A-I (Color Plate 3, Table II) is primarily responsible for the greater stability of these models. These electrostatic interactions may also be critical to the differences in function between these molecules. Color Plate 3 shows the Asp72 to Thr236 fragment of DgA-I as fitted to the apo Lp-IIIa template. Although apo Lp-III and the apo A-I proteins all have significant potential for the formation of amphipathic helices, the actual number of identities across these sequences is only 10%. Thus the substitution of apo A-I sequences into the apo Lp-IIIa template was reasonably successful, and may represent another path in the search for structural correlations among distantly related proteins.

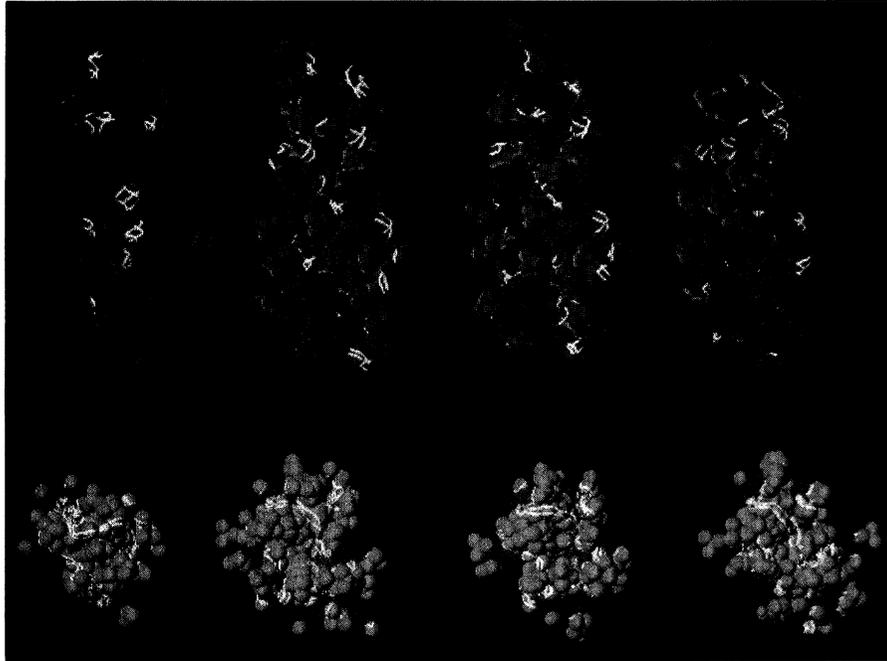
### Acknowledgments

The authors wish to thank Drs. James Chen, Edyth Malin and Gregory King for technical assistance and helpful suggestions.

### Literature Cited

1. Chapman, M.J. *Meth. Enzymol.* **1986**, *128*, 70-143.
2. Weisgraber, K.H. *Adv. Protein Chem.* **1994**, *45*, 249-302.
3. Lux, S.E.; Hirz, R.; Shrager, R.I.; Gotto, A.M. *J. Biol. Chem.* **1972**, *247*, 2598-2606.
4. Osborne, J.C., Jr.; Brewer, H.B., Jr. *Ann. N.Y. Acad. Sci.* **1980**, *348*, 104-121.
5. Brown, E.M. *Poultry Science* **1989**, *68*, 399-407.
6. Boguski, M.S.; Freeman, M.; Elshourbagy, N.A.; Taylor, J.M.; Gordon, J.I. *J. Lipid Res.* **1986**, *27*, 1011-1034.
7. McLachlan, A.D. *Nature* **1977**, *267*, 465-466.
8. Lou, C.-C.; Li, W.-H.; Moore, M.N.; Chan, L. *J. Mol. Biol.* **1986**, *187*, 325-340.
9. De Loof, H. *Ann. Biol. Clin.* **1988**, *46*, 10-15.
10. Brasseur, R. *J. Biol. Chem.* **1991**, *266*, 16120-16127.
11. Chou, P.Y.; Fasman, G. D. *Adv. Enzymol.* **1978**, *47*, 45-148.
12. Garnier, J.; Osguthorpe, D.J.; Robson, B. *J. Mol. Biol.* **1978**, *120*, 97-120.
13. Eisenberg, D. *Annu. Rev. Biochem.* **1984**, *53*, 595-623.
14. Kyte, J.; Doolittle, R. F. *J. Mol. Biol.* **1982**, *157*, 105-132.
15. Andrews, A.L.; Atkinson, D.; Barratt, M.D.; Finer, E.G.; Hauser, H.; Henry, R.; Leslie, R.B.; Owens, N.L.; Phillips, M.C.; Robertson, R.N. *Eur. J. Biochem.* **1976**, *64*, 549-563.
16. Mahley, R.W., Innerarity, T.L.; Rall, S.C. Jr.; Weisgraber, K.W. *J. Lipid Res.* **1984**, *25*, 1277-1294.
17. Segrest, J.P.; Jackson, R.L.; Morrisett, J.D.; Gotto A.M., Jr. *FEBS Lett.* **1974**, *38*, 247-253.
18. Li, W.H.; Tanimura, M.; Luo, C.C.; Datta, S.; Chan, L. *J. Lipid Res.* **1988**, *29*, 245-271.
19. Segrest, J.P.; Garber, D.W.; Brouillette, C.G.; Harvey, S.C.; Anantharamaiah, G.M. *Adv. Protein Chem.* **1994**, *45*, 303-369.

20. Smith, L.C.; Pownall, H.J.; Gotto, A.M. Jr. *Annu. Rev. Biochem.* **1978**, *47*, 751-777.
21. Glomset, J.A. *J. Lipid Res.* **1968**, *9*, 155-167.
22. Blue, M-L.; Ostapchuk, P.; Gordon, J.S.; Williams, D.L. *J. Biol. Chem.* **1982**, *257*, 11151-11159.
23. Soulagés, J.L.; Wells, M.A. *Adv. Protein Chem.* **1994**, *45*, 371-415.
24. Breiter, D.R.; Kanost, M.R.; Benning, M.M.; Wesenberg, G.; Law, J.H.; Wells, M.A.; Rayment, I.; Holden, H.M. *Biochemistry* **1991**, *30*, 603-608.
25. Wilson, C.; Wardell, M.R.; Weisgraber, K.H.; Mahley, R.W.; Agard, D.A. *Science* **1991**, *252*, 1817-1822.
26. Nolte, R.T.; Atkinson, D. *Biophys. J.* **1992**, *63*, 1221-1239.
27. National Biomedical Research Foundation, Georgetown University Medical Center, Washington DC 20007. Release 28.0 March 31, **1991**.
28. Cole K.D.; Fernando-Warnakulasuriya, G.J.P.; Boguski, M.S.; Freeman, M.; Gordon, J.L.; Clark, W.A.; Law, J.H.; Wells, M.A. *J. Biol. Chem.* **1987**, *262*, 11794-11800.
29. Vazquez, S.R.; Kuo, D.Z.; Botsitis, C.M.; Hardy, L.W.; Lew, R.A.; Humphreys, R.E. *J. Biol. Chem.* **1992**, *267*, 7406-7410.
30. Brasseur, R.; Lins, L.; Vanloo, B.; Ruyschaert, J-M.; Rosseneu, M. *Proteins* **1992**, *13*, 246-257.
31. SYBYL, Macromolecular modeling software, version 5.5, TRIPOS Associates, Inc. St. Louis, MO.
32. Weiner, S.J.; Kollman, P.A.; Case, D.A.; Singh, U.C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P.K. *J. Am. Chem. Soc.* **1984**, *106*, 765-784.
33. Weiner, S.J.; Kollman, P.A.; Nguyen, D.T.; Case, D.A. *J. Comput. Chem.* **1986**, *7*, 230-252.
34. Schiffer M.; Edmunson A.B. *Biophys. J.* **1967**, *7*, 121-125.
35. Kumosinski, T.F.; Brown, E.M.; Farrell, H.M., Jr. *J. Dairy Sci.* **1991**, *74*, 2879-2888.
36. Pinker, R.J.; Lin, L.; Rose, G.D.; Kallenbach, N.R. *Protein Science* **1993**, *2*, 1099.
37. Marcel, Y.L.; Provost, P.R.; Koa, H.; Raffai, E.; Dac, N.V.; Fruchart, J-C.; Rassart, E. *J. Biol. Chem.* **1991**, *266*, 3644-3653.
38. Banka, C.L.; Bonnet, D.J.; Black, A.S.; Smith, R.S.; Curtiss, L.K. *J. Biol. Chem.* **1991**, *266*, 23886-23892.



Color Plate 3. Energy refined models of from left to right apo lipophorin-III (residues 7 - 156), canine apolipoprotein A-I (residues 72 - 236), human apolipoprotein A-I (residues 73 - 237) and chicken apolipoprotein A-I (residues 72 - 236). Peptide backbones are represented by a double stranded ribbon. In the lateral view, only ionizable sidechains are displayed with acidic groups (Glu, Asp) in red and basic groups (Lys, Arg) in blue. In the end-on view only hydrophobic sidechains (Ala, Ile, Leu, Met, Phe, Tyr, Trp, Val) are displayed in orange.