

6072

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

Predicted Energy-Minimized α_{s1} -Casein Working Model

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

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

A previously reported three dimensional model of α_{s1} -casein (J. Dairy Sci 74: 2889, 1991) was constructed using sequence based prediction algorithms in conjunction with global experimental secondary structural information obtained from Raman Spectroscopy. This model is now energy minimized using primarily, the Kollman force field. Both the original and energy minimized structures contain a hydrophilic domain and a hydrophobic domain which are connected by a segment of α -helix. However, within the hydrophobic domain, three anti-parallel hydrophobic sheets are in different spatial orientations for the two structures. To mimic the self-association properties of the α_{s1} -casein B at low ionic strength, a tetramer model was constructed using the largest hydrophobic antiparallel sheets and the hydrophobic ion-pair, which occurs within the deletion peptide of the A variant, as interaction sites. Two tetramers were self associated via hydrophobic sites to model the octamer of α_{s1} -casein B which occurs at high ionic strengths. All energy minimized working models are in agreement with many of the biochemical, chemical and physico-chemical properties of α_{s1} -casein A, B and C.

In a previous report (16), a predicted three dimensional model of α_{s1} -casein was presented. This structure was built using sequence based secondary structure prediction algorithms, global secondary structural information from Raman Spectroscopy and molecular modeling techniques which, in part, minimized bad van der Waals contacts. The model consisted of a hydrophilic domain which contained most of the serine phosphates and a hydrophobic domain which contain two (one large and one small) stranded anti-parallel β -sheet structures with hydrophobic side chains. Both domains were connected via an extended α -helix, and another sheet structure. In addition, the model accounted for a variety of functional properties, derived biochemical, chemical and physico-

chemical results for α_{s1} -casein A, B and C. However, this structure was not energy minimized for interactions resulting from van der Waals energies, electrostatics and intra-molecular hydrogen bonding. In fact, a destabilizing energy of over two million kcal/mole could be calculated for this structure using a Kollman force field (31).

In this paper, we will present a structure refined by energy minimization techniques to further improve the original α_{s1} -casein model. This model will be compared with the original structure and discussed with respect to structural motifs present in α_{s1} -caseins of other species, biochemical cleavage information via renin, chemical modification results, and solution physico-chemical experiments. Also, to mimic the self-association of α_{s1} -casein, interaction sites on the monomer will be utilized to build energy minimized tetramer and octamer structures.

Materials and Methods

Predictions of Secondary Structures. Selection of appropriate conformational states for the individual amino acid residues was accomplished by comparing the results of sequence-based predictive techniques, primarily those of Chou and Fasman (5), Garnier et al. (12) and Cohen et al. (6,7). Assignments of secondary structure (α -helix, β -sheet or β -turn) for the amino acid sequences were made when either predicted by more than one method, or strongly predicted by one and not predicted against by the others. Such methods have previously been applied to the caseins (9,13,19). In addition, because of the large number of proline residues in the caseins, proline-based turn predictions were made using the data of Benedetti et al. (2) and Ananthanarayanan et al. (1).

Energy Minimization - Molecular Force Field. The concept, equation for and a full description of a molecular force field was given in a previous communication (17). In these calculations, a combination of the Tripos (24) and Kollman force fields (31) were employed. Both force fields used electrostatic interactions calculated from partial charges given by Kollman (31). A united atom approach with only essential hydrogens for reasonable calculation time on the computer was also used. A cutoff value of 8 Å was employed for all non-bonded interactions for both force fields. Both the BFGS (Boyden, Fletcher, Goldfarb and Shanno) and conjugate gradient techniques were employed as minimization algorithms when applicable. The Tripos force field contains fewer parameters than the Kollman force field, i.e. no 6-12 potential for specific hydrogen bonding, and is used in this study to overcome large energy barriers.

Molecular Modeling. The three dimensional structure for α_{s1} -casein was approximated using molecular modeling methods with an Evans and Sutherland

PS390¹ interactive computer graphics display driven by Sybyl (Tripos, St. Louis, MO) software on a Silicon Graphics (Mountainview, CA) DW-4D35 processor. The original structure was built with assigned ϕ and ψ angles characteristic of the respective predicted structures. All ω angles were assigned the conventional trans configuration. In addition, aperiodic structures are in the extended rather than totally random configuration. The Sybyl subroutine "SCAN" was used, on the side chains only, to adjust torsional angles and relieve bad van der Waals contacts. The individual pieces were then joined together to produce the total polypeptide model. This initial structure has been presented in a previous paper (16).

The original 3D model, which was constructed using a combination of sequence based secondary structure prediction algorithms constrained to be compliance with experimental global secondary structure determined from vibrational spectroscopy, was divided into three parts. The cleavage point was chosen to be in the middle of an extended structure to allow joining for reassembly of the total structure after minimization with known ϕ and ψ angles. Since secondary structure sequence based prediction algorithms are not informative with the respect to the type of turns, each of the three portions of the α_{s1} -casein B structure was reconstructed with differing turns and energy minimized. Each structure were tested by this method, and the model with the lowest energy was chosen for reassembly with the other pieces. Thus, by this method a larger sampling of conformational space was performed. However, this methodology does not in any way allow one to conclude that the global energy minimization structure was achieved.

The energies of the three pieces of the initial structure were calculated using a Tripos and a Kollman force field. Because of the large destabilizing van der Waals, bond stretching, due to the high proline (see Figure 1) and turn content, as well as the positive hydrogen bonding energy, each piece was first minimized with respect to total energy using a Tripos force field without electrostatics. The BFGS technique (24), which requires a large block of computer memory, was employed as the minimizing algorithm. This technique is more useful when a large number of energy barriers are suspected. The pieces were then subjected to a Tripos force field with electrostatics and finally a Kollman force field. The conjugate gradient technique was used only when the Kollman force field was employed.

The three energy minimized portions of α_{s1} -casein B were then joined together using ϕ , ψ values for an extended structure. The total structure was then energy minimized to an energy of ± 0.05 kcal/mole using a conjugate gradient algorithm and a Kollman force field (31).

The above procedure is an extension of the method of Cohen and Kuntz (38) where sequence based secondary structure prediction methods are used as a

¹Mention of brand or firm name does not constitute an endorsement by USDA over products of a similar nature not mentioned.

starting point for tertiary structure prediction. Since the choice of the sequence based algorithm is arbitrary (dubious) an added constraint of agreement between the individual or consensus algorithm and experimental global secondary results in conjunction with energy minimization techniques was utilized for sampling more of the possible conformational space.

This method may be tested with a protein whose global secondary structure and X-ray crystallographic structure is known. In addition, the protein, like casein should contain no disulfide bonds. Avian pancreatic polypeptide with 63 residues was chosen. Here, the global secondary structure results were obtained from the circular dichroism (CD) experiments of Noelken et al. (22) which showed a minimum of 80% helix under a variety of environmental conditions. The consensus prediction algorithm which agreed with the CD results showed a polyproline helix for residues 1-9, α helix in residues 13-32 and turns at residues 10 and 11. All other residues were given an extended conformation. After several energy minimization processes during which the turns were changed, the structure with the model with lowest energy, i.e. -599.4 kcal/mole, was chosen and is shown as a backbone ribboned structure in the lower portion of Figure 2. The upper portion shows the backbone of the X-ray crystal structure (1PPT) in the representation. Comparison of the backbone atoms between the predicted model X-ray structure yielded a root-mean square deviation of 3.37Å using the algorithm fit of the Sybyl program. Hence, a reasonable low resolution structure can be obtained by this methodology. It should be stressed that all structures built by these methods should be viewed as low resolution working models and not perfect structures for which the global energy minimum has been attained.

Assembly of Aggregate Structures. Aggregates were constructed using a docking procedure on the modeling system previously described. The docking procedure of this system allowed for individually manipulating the orientation of up to four molecular entities relative to one another. The desired orientations could then be frozen in space and merged into one entity for further energy minimization calculation utilizing a molecular force field. The criterion for acceptance of reasonable structures was determined by a combination of experimentally determined information and the calculation of the lowest energy for that structure. At least ten possible docking orientations were attempted, each structure was then energy minimized and assessed for the lowest energy in order to provide a reasonable sampling of conformational space.

Results and Discussion

Generation of Energy Minimized Three Dimensional Models. The caseins in general represent a unique class of proteins which are neither globular nor fibrous. They are characterized by a high content of the amino acid proline. In our initial attempts at undertaking structural motifs for casein (15,16) we were

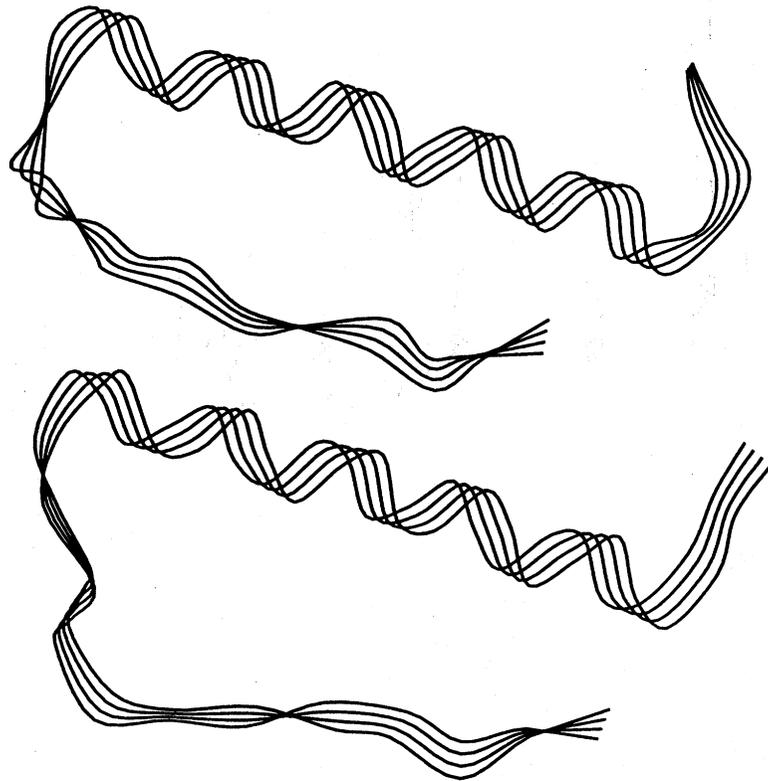


Figure 2. Ribboned backbone models of avian pancreatic polypeptide, upper: X-ray structure (1PPT); lower: predicted model.

struck by the fact that data from Raman spectroscopy predicted a high degree of turns and that there appeared to be a correlation between the percent proline and the percent of predicted turns. As pointed out by others, proline disrupts regular structure but excels at making turns (1,2,7,27). In globular proteins turns are almost always on the surface and are hydrophilic in nature. Caseins, however, have a propensity to self-associate (10,13,28) and most data point toward hydrophobic cores in the associated complexes (10,28). Thus proline-driven turns in hydrophobic areas may serve to facilitate the formation of interaction sites for hydrophobic core formation. Therefore, in the construction of casein monomers, proline turns in hydrophobic areas can appear on the monomer surface because it is anticipated these will be buried in subsequent polymer formation. In point of fact, the absence of a hydrophobic modeling term and the lack of water in these in vacuo calculations may actually be a positive element in these studies. The monomer constructed here actually could represent the monomer-within-a-polymer for the individual caseins. Exposed hydrophobic sites help to explain the nature of casein self-associations. In fact, these hydrophobic proline-based turns could well be the signature of an α_{s1} - or κ -casein, the structural motif which sets these proteins apart from both globular and fibrous proteins and even to some extent β -casein (17).

In the work that follows, it must be understood that we begin with secondary structural predictions. These predictions do not have a high degree of certitude, but they are based upon bond angles which occur in aqueous crystals for proteins. The beginning structures, like the previously proposed linear model (9,13) are modified to account for global circular dichroic and Raman data for α_s -casein, so that the assigned structures are not altogether arbitrary nor without precedent (9,13). Energy minimization of these structures could trap a less than favorable energetic state, but we have used algorithms which could avoid this problem. However, given the task at hand, de novo calculations from randomized structures would also not guarantee a native structure for a molecule the size of α_{s1} -casein. The structure presented herein thus represents a computational short-cut, and therefore is a working model, subject to change and refinement, and not a final exact structure. It is presented because it is unlikely that exact crystal structures will ever be available, and it is meant to stimulate research and discussion. In that light, the following steps were used to assemble the α_{s1} -casein model.

The refined, energy minimized model of α_{s1} -casein B was built using the computer generated structure presented in a preceding publication (16). In this latter paper, three segments were individually built, residues 1-84, 85-160 and 161-199, and then joined with appropriate ϕ , ψ angles to produce the polypeptide chain. Here, we individually energy minimized the three pieces by the following procedure. For the N-terminal piece, a destabilizing energy of over 900,000 kcal was calculated for the initial model using a Kollman force field as seen in Table I. A large portion of this energy resulted from improper van der Waals and hydrogen bonding interactions, with a smaller contribution

attributed to bond stretching energies. This large net energy is most likely due to the conformation of the six proline residues which occur in the first half of this segment of α_{s1} -casein (Figure 1). Because of these high energy values, energy minimization was performed using a Tripos force field first without electrostatic and then with electrostatic interactions. For these calculations, a BFGS algorithm was used for the minimization algorithm in order to avoid high energy wells with local minima. This combination of methods proved successful in reducing the calculated energy; results of the Tripos energy minimization with electrostatics are presented in column two of Table I. Here, the van der Waals energies are considerably lower than the initial values presented in column one of the same Table and H-bond energies are omitted by the program. Finally, this structure was refined using a Kollman force field in conjunction with a faster conjugate gradient minimization algorithm which takes H-bonding into account. The results are presented in column three of Table I, which shows a favorable total energy of -682 kcal/mole.

TABLE I. Energy calculations for the N-terminal segment of α_{s1} -casein (residues 1-84)

Energies k cal/mole	Initial	Tripos	Kollman
Bond Stretching	4223	365	42
Angle Bending	683	437	196
Torsional	440	207	246
Out of Plane Bending	53	53	21
1-4 van der Waals	206	-24	118
van der Waals	590463	171	-346
1-4 Electrostatic	780	155	905
Electrostatic	-2474	-2612	-1844
H-Bond	319688	---	-20
Total	914062	156	-682

The same procedure was applied to the other two pieces of the initial structure. The three minimized structures were then joined with the appropriate ϕ , ψ angles determined from the secondary structure sequence based prediction results presented in the preceding paper (16). This final total polypeptide chain was further energy minimized using a Kollman force field with a conjugate gradient minimizer. The result of this calculation as, seen in Table II, yield a

stabilizing energy of -2002 kcal/mole or ≈ 10 kcal/mole/residue. Such values are consistent with results obtained from the use of force fields to energy minimize structures derived from X-ray crystallography.

Refined Three Dimensional Structure of α_{s1} -Casein. The energy minimized structure, generated for α_{s1} -casein B as described above is shown in Color Plate 19 where it is displayed from carboxyl- to amino-terminal (left to right). Analysis of this structure shows the molecule to be composed (right to left) of a short hydrophilic amino-terminal portion, a segment of rather hydrophobic β -sheet, the phosphopeptide region, a short portion of α -helix connects this N-terminal portion to the very hydrophobic carboxyl-terminal domain containing extended antiparallel β -strands. For clarity the backbone without side chains is shown in Figure 3A with prolines (P) indicated, and an accompanying α -carbon chain trace stereo view (Figure 3B) is given.

TABLE II. Energy calculation for the refined α_{s1} -casein B structure

Bond Stretching Energy :	34	
Angle Bending Energy :	425	
Torsional Energy :	427	
Out of Plane Bending Energy :	16	
1-4 van der Waals Energy :	306	
van der Waals Energy :	-872	
1-4 Electrostatic Energy :	2135	
Electrostatic Energy :	-4426	
H-Bond Energy :	-47	
Total Energy :	-2002	kcal/mol

From the overall shape of the α_{s1} - model (Color Plate 19 and Figure 3), it is apparent that neither a prolate nor an oblate ellipsoid of revolution can be used to approximate its structure, as was done in the case of the β -casein refined structure (17). Indeed, a rather large degree of asymmetry of structure is observed. As noted above the hydrophilic and hydrophobic domains are joined by extended structures whose central feature is an α -helix with its pitch perpendicular to the two domains. It is speculated that this α -helix would be important for preserving the integrity of the two domains when a dynamic calculation is performed.

In the bovine casein this segment of α -helix occurs from residues 92 to 100 (Color Plate 19 and Figure 3). In the protein from all three ruminant milks, starting with residue 90 there are no amino acid substitutions in this region. For rat α_{s1} -casein a near exact or functional homology occurs for the beginning

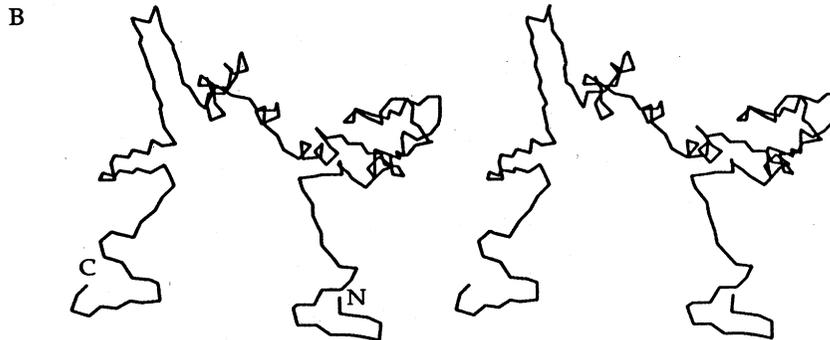
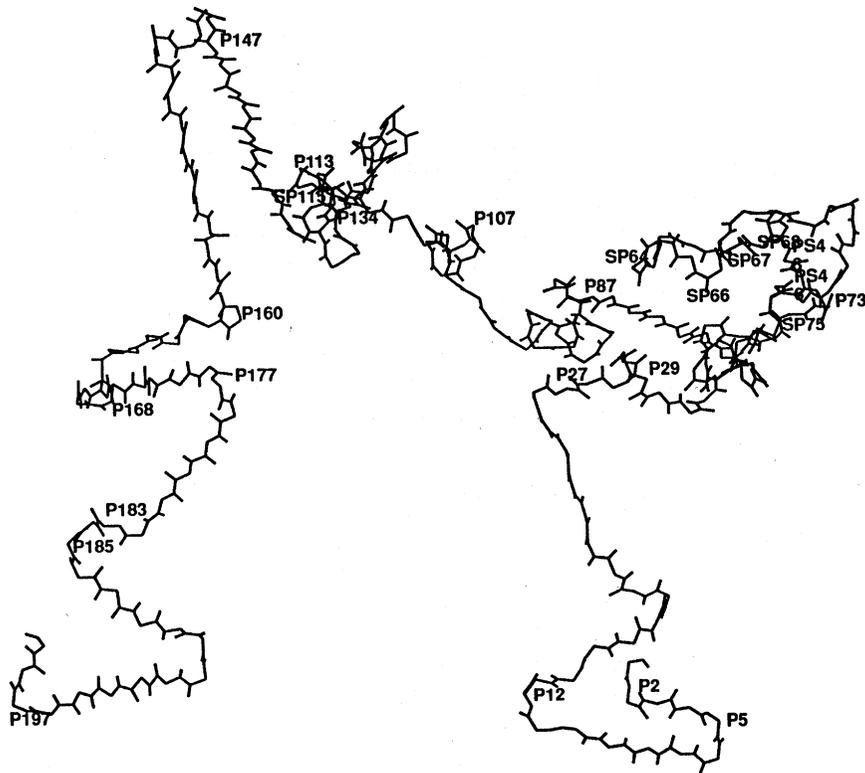
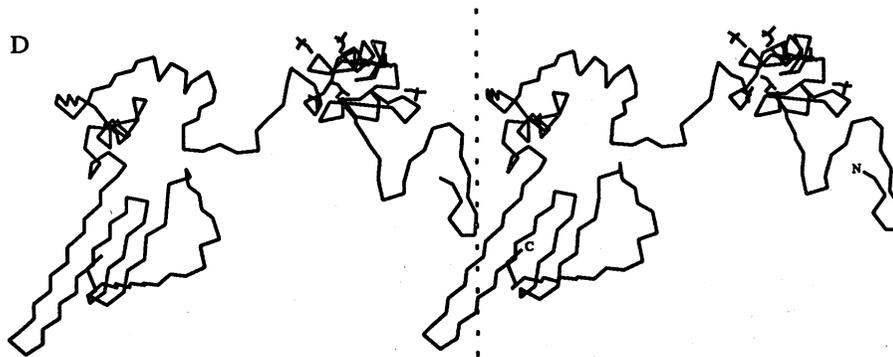
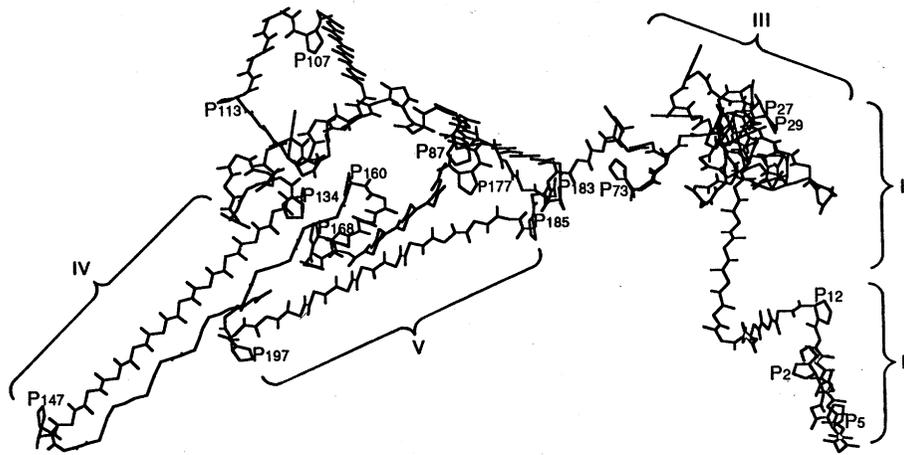


Figure 3. A) Backbone of the refined model of α_{s1} -casein, without side chains, prolines (P) indicated. B) Stereo view of the refined three dimensional molecular model of α_{s1} -casein; the N- and C-terminal ends of the molecule are labelled. C) Initial backbone structure of α_{s1} -casein B (Kumosinski et al., 16). D) Stereo chain trace of initial structure of a comparison with refined energy minimized structure. The ticked line represents the suggested stereo center, where the center of the stereo viewer should be placed.



of the α -helical region. In the rodent α_{s1} - however, there is then a large segment of repeating structure inserted at this point (Figure 1). It is interesting to speculate that this segment, which is based on the DNA-sequence, may also be helical in nature. Holt and Sawyer (13) calculate β -sheet for this extended region but in either case the rodent molecule has an even greater space between the hydrophilic and hydrophobic domains of the protein.

For bovine α_{s1} -casein, a radius of gyration (Rg) of 31 Å with a dipole moment of 3002 Debye units and a net charge of -26 can be calculated from its structure. In contrast, the refined β -casein structure yields an Rg of only 21 Å and a dipole moment of an 1825 D (17). These values demonstrate the large global differences between the α_{s1} -casein and β -casein refined structures, which may account for their differing functional properties, especially with respect to submicelle and micelle formation. The backbone structure without side chains as well as the stereo chain trace structure for the unrefined initial structure (15) is presented in Figure 3C and D, respectively. Comparison of Figures 3A and B with C and D, show major and minor differences between the refined energy minimized structure and the initial model. Although the initial and refined structures both consist of a hydrophilic and hydrophobic domain, the spatial orientation in the hydrophobic domain is quite different. A major change in the structure centering about proline 134 opens the structure dramatically. Thus the long 20 residue hydrophobic stranded antiparallel β -sheet (centered on proline 147) is opened up as is the short 8 residue piece (centered on proline 168). These are the major differences between the initial and refined structures Figure 3A, B and Figure 3C, D, respectively. This orientational difference resulted from the large portion of van der Waals energies from the proline based turns of the initial structure. In addition, the van Waals energies in the hydrophilic region due to proline and serine phosphates in the initial structure are reduced because these regions are partially changed to loops in the refined structure.

To test the accuracy of the refined structure of α_{s1} -casein B, comparisons between this model and experimental data, derived from solution studies such as Raman spectroscopy, chemical and biochemical experiments and solution physico-chemical studies on the hydrophobically induced, ionic strength dependent self-association of α_{s1} -casein, are presented.

Secondary Structure Analysis. The Ramachandran plot of ψ , ϕ , backbone dihedral angles (open circles) calculated from the refined α_{s1} -casein B structure, using the Tripos' Sybyl molecular modeling software is shown in Figure 4. Also, bounded areas of acceptable ψ , ϕ angles for β -sheet, α -helix and β - and γ -turn regions are presented with appropriate labels. Hence, by summing the number of points that are present within the limits of the particular periodic or turn structure, the global amount of α -helix, β -sheet or turn can be obtained. However, this type of analysis does not allow for the residue length of the periodic structure or its placement within the sequence structure. Hence, a visual inspection of the secondary structure by use of a chain trace or ribbon as

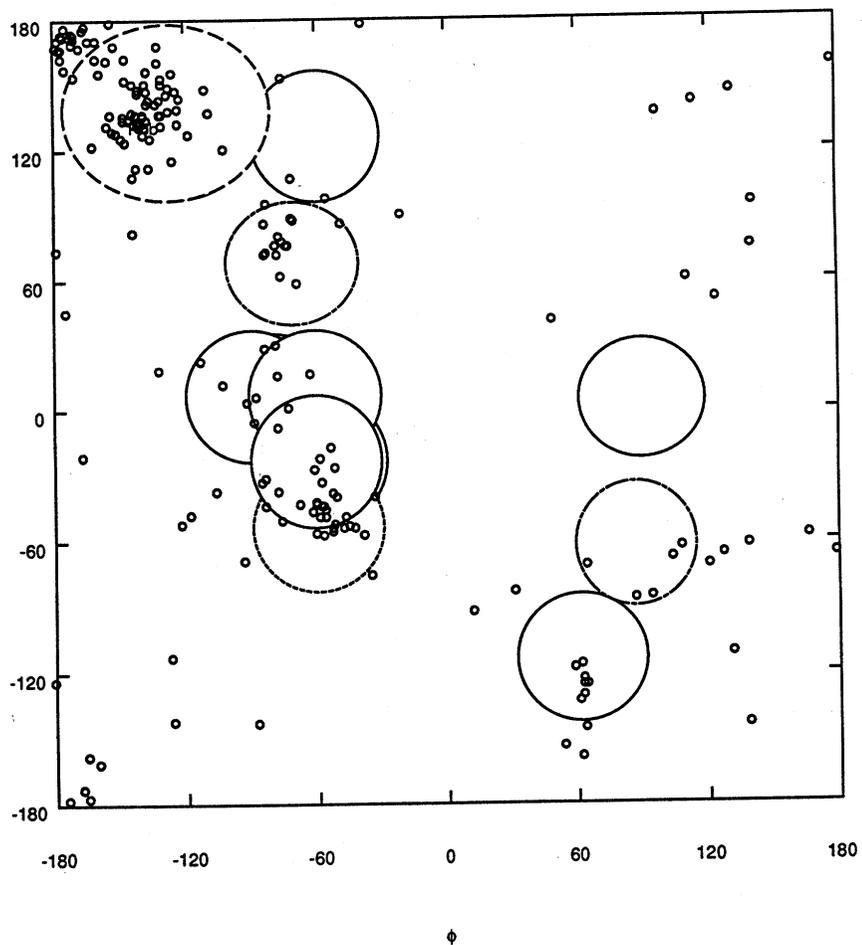


Figure 4. Ramachandran plot of ψ , ϕ angles calculated from energy minimized structure of α_{s1} -casein B. Dashed line (—) indicates area for β -sheet structure, dotted line (···) for α -helix, solid lines for β -turns, and dashed-dot line (— · — ·) γ -turns.

in Color Plate 19 should also be employed when calculating the global secondary structure of any model.

The above procedure was employed on the α_{s1} -casein B initial and refined structure and, as seen in Table III, the global secondary structural results are in good agreement with those obtained via Raman spectroscopy in D₂O. In fact, the refined structure yields values which agree more closely with the Raman results than the initial structure.

Holt and Sawyer (13) have recently used other secondary structure prediction algorithms and have theorized that α_{s1} -casein could contain up to 10% α -helix, 10% turns, 50% β -structure and 30% unordered structure. Their predictions are by their nature biased toward β -structure and so relative to the experimental Raman data (Table III) β -sheet is over predicted while the % of turns are underpredicted. However the N-terminal half of their predicted structure is in relatively good agreement with the 3D model presented in this work. In fact the major α -helical residues between residues 120-130 while somewhat frame shifted by one or two are also in good agreement. The major difference resides in our preselection of proline residues to be in turns as noted above. In the structures predicted by Holt and Sawyer (12) expansion of the % turns to include prolines and their adjacent 2 to 3 residues would increase their turn content primarily at the expense of β -sheet and bring their predictions closer to the global Raman data. Interestingly, our hydrophobic based turns, which depend upon prolines 147, 160, 168, 177 and 185 are in areas predicted by Holt and Sawyer to contain at least one residue in a β -turn conformation. It would appear then that there is overall good but not exact agreement between the algorithms used here and those of Holt and Sawyer (13). One major difference is for the rodent protein with its large insertion (Figure 2) where our algorithms favor α -helix while that used by Holt and Sawyer (13) favors β -sheet.

TABLE III. Comparison of the initial and the final secondary structures of α_{s1} -casein with spectroscopic data

Sample		% Helix	% β -Structure	% Turns	% unspec
α_{s1} -Casein	Raman ¹	8 - 13	18 - 20	29 - 35	33 - 40
	Initial	15	22	45	18
	Refined	8	18	34	40

¹ Reference 4.

Chemistry of α_{s1} -Casein and the Refined Model

Sites of Phosphorylation in α_{s1} -casein. α_{s1} -Casein B is a single polypeptide chain of 199 amino acid residues with a molecular weight of 23,619 (20). The α_{s1} -B molecule contains eight phosphate residues, all in the form of serine monophosphate. Seven of these phosphoserine residues are clustered in an acidic portion of the molecule bounded by residues 43 and 84 (the second fifth of the molecule from the amino-terminal end). This highly acidic segment contains 12 carboxylic acid groups as well as seven of the phosphoserines. The model shows all seven of the phosphate residues in this cluster to be located on β -turns which is compatible with known phosphorylated residues in crystallized proteins (27). The folding of the chain brings these phosphate residues into close proximity. This cluster forms a highly hydrophilic domain on the right shoulder of the molecule (Figure 3) and is bounded by prolines 29 and 87.

The α_{s1} -A Deletion and Chymosin Cleavage Sites. The rare α_{s1} -casein A genetic variant exhibits self association reactions which are highly temperature dependent (10,11,13). The A variant is the result of the sequential deletion of 13 amino acid residues between residues 13 and 27; the majority of these deleted amino acids are apolar (11) but Glu 14, Glu 18 and Arg 22 are also deleted (11). The deleted segment encompasses a region predicted to be in a β -sheet. This sheet provides a spacer-arm between the hydrophilic amino-terminal region (five o'clock in Figure 3A) and the phosphopeptide region. Deletion of this spacer-arm brings the hydrophilic N-terminal section (Figure 3A) closer to the phosphate rich shoulder. In addition, the Phe-Phe bond (residues 23 and 24) is deleted in the A variant; this represents a major chymosin cleavage site (15,19) and its absence may account for the poor quality products prepared from α_{s1} -casein A milks (28). Additional chymosin cleavage sites are located between residues 32-33, 149-150, 169-170 and 179-180 (19,21). The site between residues 32 and 33 is relatively exposed thus facilitating enzymatic attack. The last three sites, 149-150, 169-170 and 179-180 appear to be less exposed and could be involved in hydrophobic sheet-sheet interactions.

Hydrophobic Interactions. For α_{s1} -casein, the high degree of hydrophobicity exhibited by the carboxyl terminal half of the molecule (residues 100 to 199) may be responsible, in part, for the pronounced self-association of the α_{s1} -casein monomer in aqueous solution (28,29). This self-association approaches a limiting size (\sim tetramer) under conditions of low ionic strength; the highly charged phosphopeptide region can readily account for this phenomenon through charge repulsions. At elevated ionic strength ($> 0.1M$) the polymer size increases to an octamer, and at ionic strengths > 0.5 , α_{s1} -casein is salted out of solution at 37°C. The hydrophobic C-terminal domain contains two segments of extended β -sheets (residues 134 to 160 and 163 to 178) and one smaller extended hydrophobic segment from 180 to 188, found to the left in Figure 3A. These segments are directed by prolines 134,

147, 160, 168, 177 and 185. The crystal structure for insulin dimer (24), each monomer contains a similar pair of extended β -strands at the monomer-monomer interface. These extended β -strands may lend stability to casein polymers through sheet-sheet interactions. Of the six noted proline residues 134, 160, 168 and 185 are conserved across the four species shown in Figure 1. In addition proline 147 and 177 are compensated for in the rat protein by proline residues at $n+4$ and $n-1$ positions respectively, so that a formal structural homology for hydrophobic interactions occurs in all four species, the differences being in the length of the three hydrophobic segments. For bovine α_{s1} -caseins tyrosine residues play an important role in interactions with κ -casein; nitration of tyrosines of α_{s1} -casein leads to decreased stability in reconstituted micellar structures (32). It is interesting to note the tyrosines at 159, 166, 173 are conserved across species as are the two tryptophan residues. Residues 150 through 185 contain 13 hydrophobic residues which are exactly or functionally conserved (Figure 1). Thus the hydrophobic nature of these proline directed turns appear to be functionally preserved in most species. Such extended sheets are also predicted for κ -casein (15) so that sheet-sheet interactions may play an important role in casein micelle formation.

In α_{s1} -casein C, Glu 192 of the B variant is replaced by Gly through a point mutation (17). The association properties of α_{s1} -C are changed relative to α_{s1} -B by this replacement (21); the C variant has stronger associative properties. The segment of the bovine α_{s1} -molecule from residues 188 to 192 represents a hydrophilic turn centered on a functionally conserved amide (N/D) at positions 190 (Figure 1). This hydrophilic turn is followed, in ruminants, by a lysine residue at 193. The conversion of 192 to a glutamic residue in the B variant may alter the charge distribution here perhaps leading to weaker associative properties.

Correlation with Physico-Chemical Studies. To date, no indepth small angle X-ray or neutron scattering studies have been reported on any of the variants of α_{s1} -casein, especially, at low temperatures and ionic strengths where α_{s1} -caseins disaggregate. Hence, no correlation between an experimental radius of gyration and a value calculated from the refined structure is possible. However, there are light scattering studies on the B and C variants of α_{s1} -casein from which a stoichiometry of the α_{s1} -casein self-association maybe obtained under a variety of temperatures and ionic strengths. From the results, it was concluded that α_{s1} -casein undergoes a concentration dependent reversible association from monomer to dimer then tetramer, hexamer, octamer and even higher if the ionic strength is increased (28,29). To test the refined three dimensional structure presented in this paper, we attempted to construct energy minimized dimer, tetramer and octamer structures using primarily hydrophobic sites.

The first step was to create a dimer using the large antiparallel stranded β -sheet i.e., residues 136-158 as the interaction site. The side chains are predominately hydrophobic and the hydrogen bonding of the antiparallel sheet

secondary structure lends rigidity to this site. After energy minimization, a dimer could easily be formed if two of these stranded sheets are docked in an antiparallel fashion (Figure 5A). Such an asymmetric arrangement could minimize the dipole-dipole interactions of the backbones while allowing the hydrophobic side chains to freely interact. In fact, this minimized structure (Figure 5A) maintains a net stabilizing energy of -520 kcal/mole, i.e., $-520 = E_2 - 2 \cdot E_m$ where E_2 is the energy of the dimer and E_m is the monomer energy. Here an interesting prediction on the self association of the ovine α_{s1} -protein can be made; residues 141 through 148 are deleted. This would considerably shorten the length and symmetry of this extended sheet as turn centered at proline 147 is replaced by proline 160 while prolines 168, 177 and 185 are conserved. Thus ovine α_{s1} -casein should have weaker self associations than goat or cow. For the rodent α_{s1} -casein a proline follows 5 residues later (rat proline 230) making a better homology with cow and goat for this segment.

Another possible interaction site for hydrophobic dimerization resides in the deletion peptide of α_{s1} -casein A (i.e. the peptide, residues 14 to 26 inclusive, which is deleted from α_{s1} -casein B to form α_{s1} -casein A). Closer inspection of the residues of this peptide show a β -sheet secondary structure with hydrophobic as well as acidic and basic side chains. Thus, by docking two molecules in an antiparallel fashion, a hydrophobically stabilized inter-molecular ion pair between arginine 22, of one chain and glutamic 18 of another chain can be formed upon the construction of a dimer. The dimer was then energy minimized and is shown in Figure 5B. The formation of this hydrophobic ion pair has been used by several investigators to explain the difference in the calcium-induced solubility and colloidal stability between α_{s1} -caseins B and A (11,14). Small-angle X-ray scattering of micelles reconstituted from whole caseins containing α_{s1} -B and A show large differences with respect to submicellar packing density within the micelle structure i.e. 3 to 1 verses 6 to 1 for B and A, respectively. This difference in packing density may be a result of destructive interference in the scattered intensity due to an asymmetric structure with a center of inversion (23). It was speculated that this asymmetric structure was due to the formation of a hydrophobically stabilized intermolecular ion pair in α_{s1} -B at this deletion peptide site (21), and that the α_{s1} -A molecule, lacking this segment behaves mole like a β -casein in its properties.

A tetramer of α_{s1} -casein can be modeled starting with the dimer formed by the intermolecular hydrophobic ion pair (IPr in, Figure 5B). A molecule of α_{s1} -casein is then added to each side of this structure via the hydrophobic (Hb) sheet-sheet interaction shown in Figure 5A. Such an energy minimized tetrameric structure is presented in Figure 6A. This tetramer structure is highly asymmetric and also contains two possible hydrophobic ion pair sites at each end of the molecule. Such sites at either end of the tetramer could lead to further aggregation resulting in a rod with a large axial ratio and dipole moment. It is noteworthy that the hydrophobic antiparallel stranded sheets of residues 163-174 could not be docked with the same site one from another

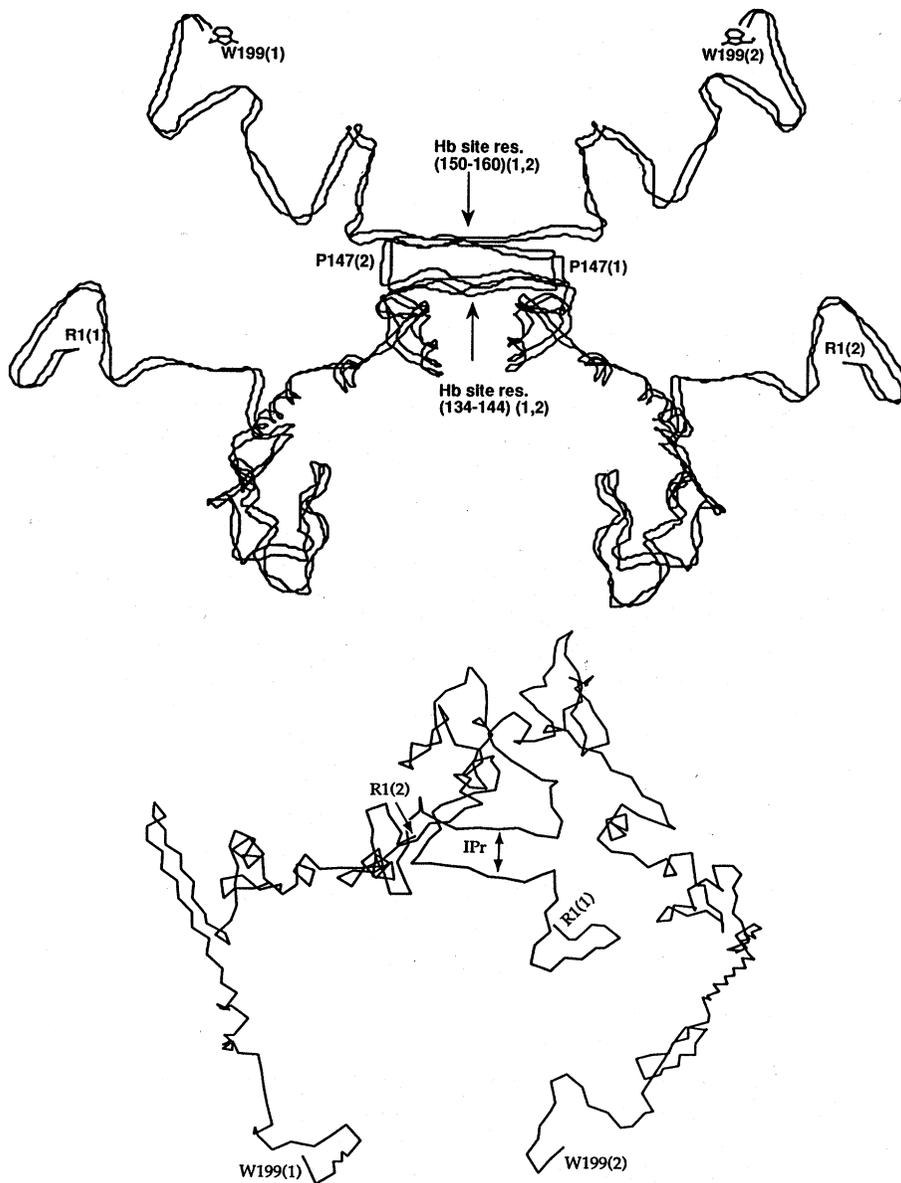
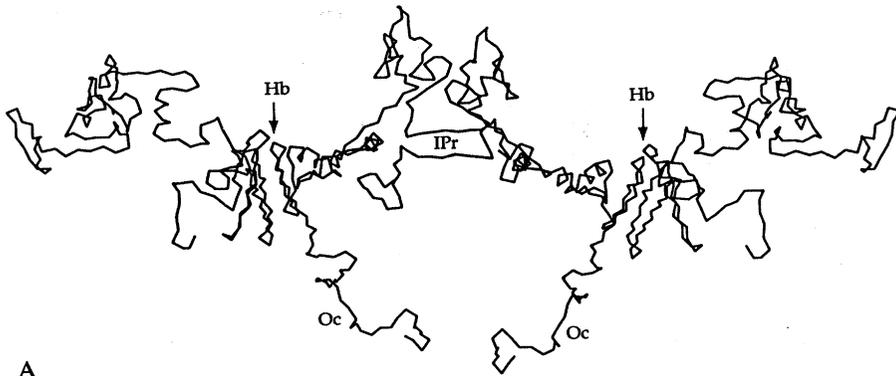
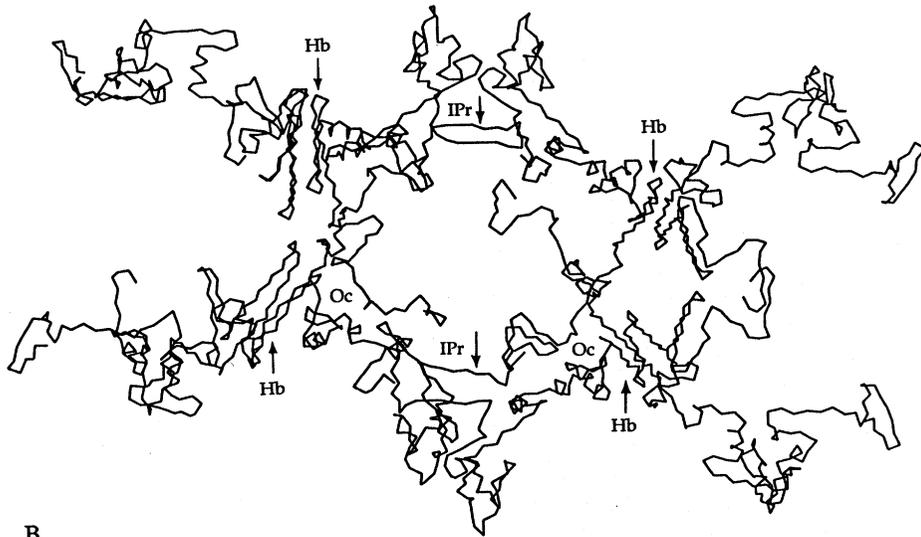


Figure 5. A) α -Carbon chain trace of backbone without side chains of hydrophobic (Hb) stabilized antiparallel sheet dimer; the large sheets centering on proline 147 are docked. B) Backbone structure of hydrophobic ion-pair dimer (IPr) from α_{s1} -casein B; this area contains the α_{s1} -A deletion peptide and centers about residues 14 to 25 in each molecule. Key for labels: R1(1), W 199(1) represent N and C terminals of molecule 1 (arginine 1 and tryptophan 199 of molecule 1); the 2 in parenthesis refers to the same residues of molecule 2. Both dimeric structures energy minimized to -10 kcal/mole/residue.



A



B

Figure 6. A) α -Carbon chain trace of α_{s1} -casein B tetramer resulting from the docking of two molecules (left and right sides) of α_{s1} -through the hydrophobic (Hb) scheme shown in 5A with the IPr (ion pair) dimer scheme given in Figure 5B. The Oc refers to potential sites for octamer formation. B) α -Carbon chain trace of octamer showing Hb, IPr and Oc interaction sites. All structures were energy minimized to -10 kcal/mole/residue.

tetramer structure due to large steric factors. Such a structure could not possibly be made without major changes in the α_{s1} -casein B model. However, a more plausible site for octamer formation via simple hydrophobic interactions can be constructed from use of the hydrophobic side chains centered between prolines 177 and 185 which are located on the lower side of the asymmetric tetramer structure (Figure 6A, Oc) and are solvent accessible. With this in mind, two tetramers were docked in an antiparallel fashion with a center of inversion using these hydrophobic side chains as interaction sites. The energy minimized octamer structure is shown in Figure 6 B and C. This model yielded a favorable energy of approximately -10 kcal/mole/residue. Such an octamer structure would still allow for water to flow through part of the polypeptide chain yielding an apparently high hydrodynamic hydration value. It would also be stable in solution since the hydrophobic side chains are predominately in the center of the model and all eight hydrophilic domains are solvent accessible, i.e., two on the upper and lower center part and two at either end of the structure (see Figure 6 B and C).

A very interesting feature of all the α_{s1} -caseins is the preservation of the C-terminal tryptophan. Ribabeau-Dumas and Garnier (25) showed that carboxypeptidase A could quantitatively remove the C-terminal tryptophan of α_{s1} -casein alone, and in native and reconstituted micelles. This was interpreted as a demonstration of the open network of the casein micelles which allowed penetration of the protease into the micelle. This is in accord with the model shown in Figure 3 A, B, C, where the C-terminal tryptophan is extended in space at the left side of the model. Thus although residues 134 to 185 participate in hydrophobic interactions a hydrophilic turn then intervenes and the C-terminal tryptophan can still be exposed in monomeric and polymeric structures. The tryptophan is thus available to digestion with carboxypeptidase A.

Finally, two of the small hydrophobic antiparallel stranded sheets of residues 163-174 are located on either end of the octamer structure between each hydrophilic domain and are therefore solvent accessible. The residues have the possibility of interaction with one or both κ -casein hydrophobic antiparallel stranded sheets even in the presence of α_{s1} -self-associations. As discussed above, tyrosines 159 and 166, proline 160 and tryptophan 164 are conserved, perhaps for relatively specific chain-chain interactions with κ -casein. Such interactions may be important in all species for micelle formation.

It is hoped that with the presentation of these working models, investigators will be inspired to perform detailed small angle scattering and other physical and biochemical experiments to ascertain the validity of these aggregate structures.

Concluding Remarks

In this paper, we have presented an energy minimized predicted three dimensional structure of α_{s1} -casein using a combination of secondary structure

sequence based prediction algorithms, global secondary structural results from Raman spectroscopy and molecular modeling techniques for energy minimization. This structure is in agreement with biochemical cleavage results using carboxy peptidase and chymosin on α_{s1} -casein B. It is also in agreement with other experimentally derived results from solution physico-chemical experiments and provides a molecular basis for the self-associations of α_{s1} -casein. However, this structure should be viewed as a working model with the ability to be changed as more precise experiments are performed to ascertain the validity and predictability of this three dimensional structure. In future studies, molecular dynamics calculations will be performed on this and the aggregate structure to test its stability when a kinetic energy equivalent to a bulk temperature is applied. In addition, it may be possible in the future to ascertain how the α_{s1} -casein B molecule specifically interacts with κ -casein to produce a synthetic submicelle structure of four α_{s1} -casein B molecules to one κ -casein molecule, the low weight-ratio complexes observed experimentally (9,13,26) in reconstitution experiments from purified caseins.

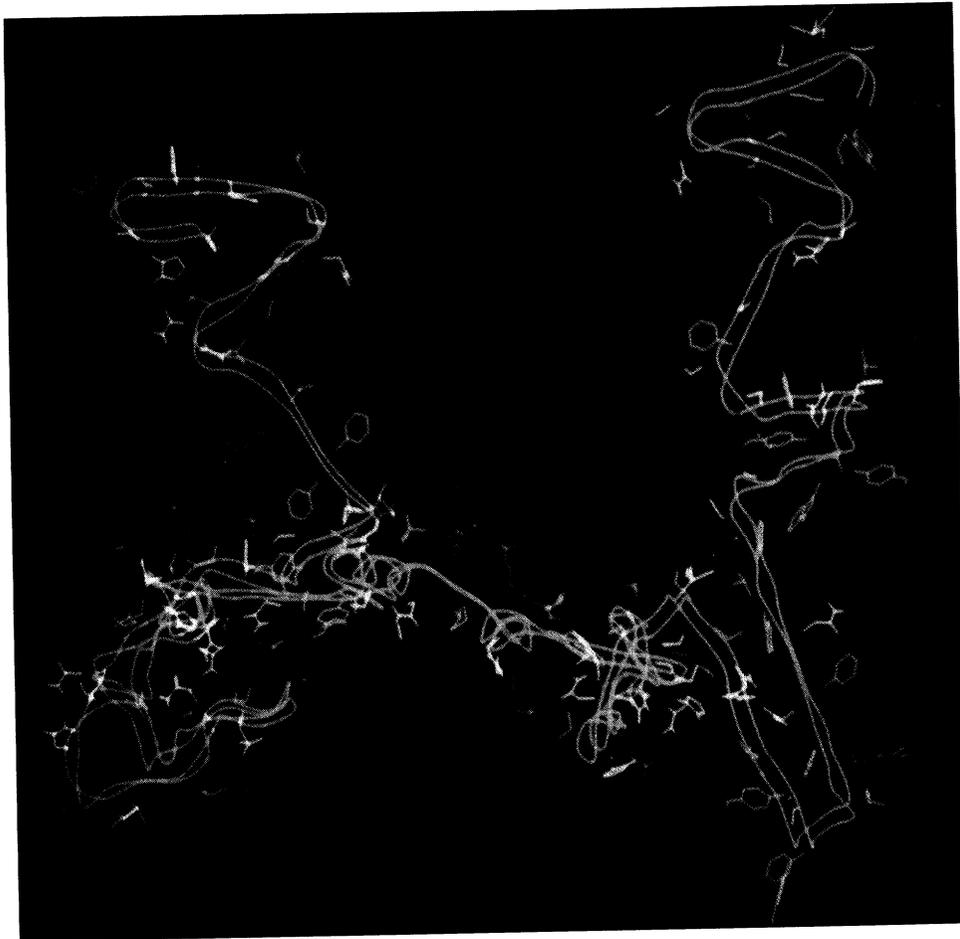
Literature Cited

- 1 Ananthanarayanan, V.S.; Brahmachari, S.K.; Pattabiraman, N. *Arch. Biochem. Biophys.* 1984, 232, 482.
- 2 Benedetti, E.; Bavoso, A.; Blasio, B.D.; Pavone, V.; Pedone, C.; Toniolo, C.; Bonora, G.M. *Biopolymers* 1983, 22, 305.
- 3 Byler, D.M.; Farrell, H.M., Jr. *J. Dairy Sci.* 1989, 72, 1719.
- 4 Byler, D.M.; Farrell, H.M., Jr.; Susi, H. *J. Dairy Sci.* 1988, 71, 2622.
- 5 Chou, P.Y.; Fasman, G.D. *Adv. Enzymology* 1978, 41, 45.
- 6 Cohen, F.E.; Abarbanel, R.M.; Kuntz, I.D.; Fletterick, R.J. *Biochemistry* 1983, 22, 4894.
- 7 Cohen, F.E.; Abarbanel, R.M.; Kuntz, I.D.; Fletterick, R.J. *Biochemistry* 1986, 25, 266.
- 8 Cohen, F.E.; Kuntz, I.D. *In Prediction of Protein Structure and the Principles of Protein Conformation*, Fasman, J.D., ed., Plenum Press, New York, 1989, p 617.
- 9 Creamer, L.K.; Richardson, T.; Parry, D.A.D.; *Arch. Biochem. Biophys.* 1981, 211, 689.
- 10 Farrell, H.M., Jr. *In Fundamentals of Dairy Chemistry*, N. Wong, Ed.; Physical Equilibria: Proteins, 3rd Edition. Van Nostrand Reinhold, New York, NY, 1988, p 461.
- 11 Farrell, H.M., Jr.; Kumosinski, T.F.; Pulaski, P.; Thompson, M.P. *Arch. Biochem. Biophys.* 1988, 265, 146.
- 12 Garnier, J.; Osguthorpe, D.J.; Robson, B.; *J. Mol. Biol.* 1978, 120, 97.
- 13 Holt, C.; Sawyer, L. *Protein Engineering* 1993, (in press).
- 14 Kumosinski, T.F.; Farrell, H.M. Jr. *J. Protein Chem.* 1991, 10, 3.

- 15 Kumosinski, T.F.; Brown, E.M.; Farrell, H.M. Jr. *J. Dairy Sci.* **1991**, *74*, 2879.
- 16 Kumosinski, T.F.; Brown, E.M.; Farrell, H.M. Jr. *J. Dairy Sci.* **1991**, *74*, 2889.
- 17 Kumosinski, T.F.; Brown, E.M.; Farrell, H.M. Jr. *J. Dairy Sci.* **1993**, *76*, 931.
- 18 Lawrence, R.C.; Creamer, L.K.; Giles, J. *J. Dairy Sci.* **1987**, *70*, 1748.
- 19 Loucheux-Lefebvre, M.H.; Aubert, J.P.; Jolles, P.; *Biophys. J.* **1978**, *23*, 323.
- 20 Mercier, J.C.; Grosclaude, F.; Ribadeau-Dumas, B. *Eur. J. Biochem.* **1971**, *23*, 41.
- 21 Mulvihill, D.M.; Fox, P.F.; *J. Dairy Research* **1979**, *46*, 641.
- 22 Noelkin, M.E.; Change, P.J.; Kimmel, J.R. *Biochemistry*, **1980**, *19*, 838.
- 23 Pessen, H.; Kumosinski, T.F.; Farrell, H.M. Jr.; Brumberger, H.; *Arch. Biochem. Biophys.* **1991**, *284*, 133.
- 24 Press, W.H.; Flannery, B.P.; Teukolsky, S.A.; Vetterling, W.T.; Numerical Recipes, The Art of Scientific Computing, Cambridge University Press, 1988, pp 301-327.
- 25 Ribadeau-Dumas, B.; Garnier, J. *J. Dairy Res.* **1970**, *37*, 269.
- 26 Richardson, J.S. *Advances Protein Chem.* **1981**, *34*, 167.
- 27 Rose, G.D.; Gierasch, L.M.; Smith, J.A. *Adv. Prot. Chem.* **1985**, *37*, 1.
- 28 Schmidt, D.G. *In Developments in Dairy Chemistry -1.*, P.F. Fox, Ed., Applied Science Pub. Ltd., Essex, England, 1982, p 61.
- 29 Swaisgood, H.E.; Timasheff, S.N. *Arch. Biochem. Biophys.* **1968**, *125*, 344.
- 30 Thompson, M.P.; Gordon, W.G.; Boswell, R.T.; Farrell, H.M. Jr. *J. Dairy Sci.* **1969**, *52*, 1166.
- 31 Weiner, S.J.; Kollman, P.A.; Nguyen, D.T.; Cose, D.A. *J. Comput. Chem.* **1986**, *7*, 230.
- 32 Woychik, J.H.; Wondolowski, M.V. *J. Dairy Sci.* **1975**, *52*, 1669.

RECEIVED July 7, 1994

Reprinted from ACS Symposium Series No. 576
Molecular Modeling: From Virtual Tools to Real Problems
Thomas F. Kumosinski and Michael N. Liebman, Editors
Published 1994 by the American Chemical Society



Color Plate 19. Energy minimized three dimensional molecular model of α_{s1} -casein. The peptide backbone has been replaced by a double yellow ribbon. Neutral side chains are colored cyan, hydrophobic side chains green, acidic side chains red, and basic side chains purple.