

(19) World Intellectual Property Organization  
International Bureau



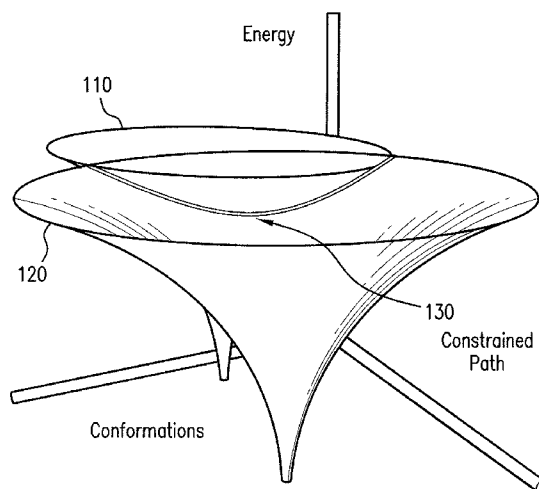
(43) International Publication Date  
6 December 2007 (06.12.2007)

PCT

(10) International Publication Number  
WO 2007/140061 A2

- (51) International Patent Classification:  
A61K 39/02 (2006.01)
- (21) International Application Number:  
PCT/US2007/067639
- (22) International Filing Date: 27 April 2007 (27.04.2007)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
60/808,167 23 May 2006 (23.05.2006) US
- (71) Applicant (for all designated States except US): **THE RESEARCH FOUNDATION OF STATE UNIVERSITY OF NEW YORK** [US/US]; Office Of Technology Licensing and Industry Relations - N5002, Frank Melville, Jr. Memorial Library, Stony Brook, NY 11794-3369 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **FORTMANN, Charles, Michael** [US/US]; 1595 Yarrow Circle, Bellport, NY 11713 (US). **KANG, Yeona** [KR/US]; 1403 Townhouse Drive, Coram, NY 11727 (US).
- (74) Agent: **RAGUSA, Paul, A.**; Baker Botts L.L.P., 30 Rockefeller Plaza, New York, NY 10112-4498 (US).
- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:  
— without international search report and to be republished upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: A METHOD FOR DETERMINING AND PREDICTING PROTEIN AUTONOMOUS FOLDING



(57) Abstract: Techniques for determining an equilibrium structure of a protein in a predetermined environment, the protein having Ramachandran angles and a known denatured structure, are disclosed. In a preferred embodiment, a method is presented which involves determining a maximum RMS volume of the known denatured structure of the protein and calculating at least one force on the protein in its current structure in the predetermined environment. The net torque resulting from the at least one force for each of the Ramachandran angles of the protein is then determined. Then at least one section of the protein structure on a side of a Ramachandran angle with greatest torque is rotated to form a new structure. A new RMS volume for the new structure is then calculated, and the method is repeated using the new structure. The method ceases when the new RMS volume of the new protein structure is not less than the RMS volume of the starting structure.

WO 2007/140061 A2

A METHOD FOR DETERMINING AND PREDICTING PROTEIN  
AUTONOMOUS FOLDING

5

## SPECIFICATION

This application claims priority from United States Provisional  
Application No. 60/808,167 filed May 23, 2006, entitled "Protein Autonomous  
Folding Units, Folding Dynamics and Structure," which is hereby incorporated by  
10 reference in its entirety.

A portion of the disclosure of this patent document contains material  
which is subject to copyright protection. The copyright owner has no objection to the  
facsimile reproduction by anyone of any portion of the patent document, as it appears  
in any patent granted from the present application or in the Patent and Trademark  
15 Office file or records available to the public, but otherwise reserves all copyright  
rights whatsoever.

BACKGROUND

The present invention relates to techniques for determining or  
predicting the secondary structure of a polypeptide, a protein, or an autonomous  
20 folding region thereof.

The secondary and higher-order structures of a protein dictate its  
function and biological activity, and are therefore critical to many fundamental life  
processes. Experimentally, such structures are often obtained through crystallography  
based on the assumption that the most stable conformation of a protein in solution is  
25 approximately the same as when it is crystallized. However, crystallography of a  
majority of proteins is currently unfeasible, presumably because chain flexibility and  
complex interactions existent within a protein molecule prevent it from forming  
crystals from solution.

Computational methods provide an alternative tool to explore protein  
30 structure, and are particularly valuable when a protein structures are inaccessible by  
crystallography or other experimental techniques. One computational approach to the  
determination of protein structure is based on energy minimization. The energy  
minimization approach, however, requires that the global (or molecular) free energy  
change to be mathematically describable in terms of local atomic bonding energies,

and therefore cannot always provide an accurate description of the most stable protein secondary structure at the free energy minimum.

Furthermore, energy minimization methods do not address response speed and axis of rotation limitations, and as a result the models generate  
5 unrealistically-quick onsets of protein folding and/or volume change. Another approach, of which Molecular Dynamics is a typical example, employs explicit Newtonian motion functions to provide a time trajectory of a molecular system evolution. *See* Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., and Haak, J.R., *J. Chem. Phys.* **81**, 8, 3684, (1984), and Karplus, Martin and  
10 McCammon, J. Andrew, *Nat. Struc. Biol.* **9**, 9, 646 (2002).

This approach is severely limited by the small time steps needed to make it physically meaningful, as well as by the complex mathematical integration schemes needed to compute the motion functions. Therefore, it is best suited for exploration of molecular phenomena occurring over a short timeframe or for small  
15 size molecules, but is of little use for molecular phenomena occurring over a longer timeframe, such as protein folding. *See* Mirny, Leonid and Shakhnivich, Eugene, *Annu. Rev. Biomol. Sturuct.* **30**, 361-96 (2001)).

Accordingly, there exists a need for a simulation method that incorporates the random forces imposed by the surrounding media and the  
20 relationship of their temporal dimension to the internal forces of a protein molecule, which method can significantly reduce the time required for a protein folding simulation, so that it becomes feasible for regular desktop computers to handle such a task.

### SUMMARY

25 The present invention provides a means to determine or predict the secondary structure of a protein, a polypeptide, or an autonomous folding region thereof. This invention allows the computational time of a task as complex as protein folding simulation to be reduced to the order of minutes on regular desktop computers, as opposed to months on supercomputers typically required by other  
30 traditional simulation methods. This acceleration is made possible by a detailed analysis of the internal driving forces, the temporal interconnection between intramolecular forces and external random forces, and the corresponding computer

codes that together describe and model the cooperative motion within a protein molecule that leads to protein folding.

Specifically, the present invention provides a method comprising (i) determining a mode of protein folding in which one part of a protein molecule moves relative to another part by a rotation around selected Ramachandran (dihedral) angles; (ii) determining the speed of protein folding in said mode under one or more internal force fields and external environmental factors such as temperature and solvent viscosity; and/or (iii) a computer code implementing such speed of motion to achieve the accelerated computation of protein folding.

The present invention has the following objectives:

To enable a physical analysis method and a computer source code that together provide a detailed description of protein and/or polypeptide and/or substructures or autonomous folding regions of a larger protein through a cooperative physical drift and diffusion model via dihedral angle change,

wherein, in certain embodiments, said physical analysis method and said computer source code employ a constant mobility or an energy coupled mobility;

and/or wherein, in certain embodiments, said physical analysis method and computer source code determine a solvent induced displacement (diffusion) at one, many, or every time step iteration used by said analysis or in said computer source code;

and/or wherein, in certain embodiments, said physical analysis method and said computer source code use calculated electrostatic force as a function of local dielectric constants to calculate a motive force;

and/or wherein, in certain embodiments, said physical analysis method employs a physical representation of forces consistent with water surface tension and bubble envelope touching force conveyance;

and/or wherein, in certain embodiments, said computer source code represents said bubble envelope touching force conveyance as a distance-force map, for example, a square distance-force map;

and/or wherein, in certain embodiments, said physical analysis method is based upon an extension of the hydrophobic interaction between two hydrophobic amino acid residues to a longer distance based upon the density and size of the equilibrium bubble concentration found in the solvent;

and/or wherein, in certain embodiments, said computer source code addresses said physical hydrophobic force extension by a two-tier square well force map and/or a multi-tier series of square-well force maps, or other approximation of said bubble extension of said hydrophobic force;

5 and/or wherein, in certain embodiments, said physical analysis method comprises determining an interaction of two or more hydrophobic and/or partially hydrophobic residues and/or regions of a protein and/or polypeptide that is predicated upon the existence of a bubble and/or an extension of the existent bubble that is dependent on the electric field strength or the distribution pattern of charged residues  
10 located between said two hydrophobic and/or partially hydrophobic residues;

and/or wherein, in certain embodiments, a computer code simulates said physical analysis method comprising determination of an interaction of two or more hydrophobic or partially hydrophobic residues and/or regions of a protein and/or polypeptide that is predicated upon the existence of a bubble and/or an extension of an  
15 existent bubble that is dependent on the electric field strength or the distribution pattern of charged residues located between in between said two hydrophobic and/or partially hydrophobic residues where said bubble or extension of said bubble is predicated upon the surface tension and/or surface energy of the environment;

and/or wherein, in certain embodiments, said physical analysis method  
20 comprises determining diffusion of one region of a protein, and/or polypeptide relative to another region utilizing a speed of motion dependent upon a mobility, the strength of the motive forces present, the viscosity of the solvent and/or changes in global energy terms such as but not limited to the global entropy;

and/or wherein, in certain embodiments, said computational method  
25 representing said physical diffusion of one region of a protein, and/or polypeptide relative to another region by a speed of motion dependent upon a mobility, the strength of the motive forces present, the viscosity of the solvent and changes in global energy terms such as but not limited to the global entropy, comprises performing a trial move to determine whether it increases or decreases a global energy  
30 term, restoring the system to the state prior to the trial move, and then moving the system again with a mobility that is corrected for the projected change in said global energy term.

The accompanying drawings, which are incorporated and constitute part of this disclosure, illustrate preferred embodiments of the invention and serve to explain the principles of the invention.

### BRIEF DESCRIPTION OF THE FIGURES

5 Fig. 1 is an illustrative embodiment of a representation of two internal energy types (shown as cones) of a protein as a function of its conformation (shape). The energy gradient (or force) is found along the curve formed by the intersection of the two energy surfaces.

10 Fig. 2 is a flowchart illustrating an embodiment of an exemplary simulation process employed to model protein folding and secondary structure formation in accordance with an aspect of the present invention.

Fig. 3 is an illustrative diagram of the calculation of torque about a backbone carbon pair created by forces acting on a protein.

15 Fig. 4 is an illustrative diagram of a protein conformation change through rotation around a Ramachandran angle about an alpha carbon backbone bond pair induced by the resolved torque operating on the bond that may be implemented in the flowchart of Fig. 2.

20 Fig. 5 is an illustrative diagram of the effect of thermal induced Ramachandran angle change on protein density useful for calculating the force equivalent of thermal energy.

25 Fig. 6 is a representation of the simulated results of Ubiquitin folding produced by one embodiment of the invention: the linear Ubiquitin as the starting structure in (A); a constant mobility, i.e.,  $\mu = D/kT$ , is used for all conformational changes of the protein, in (B); a greater mobility (2.5 fold relative to the  $\mu$  value in case B) is used for contraction while a smaller mobility (0.71 fold relative to the  $\mu$  value in case B) is used for expansion, in (C).

30 Fig. 7 is an illustrative diagram of the equilibrium probability of finding a water vapor bubble per unit volume and/or per number of water molecules as a function of bubble radius ( $2 * \text{distance}$ ) under different pressures in the absence or in the presence of electric fields according to one embodiment of the invention.

Fig. 8 is an illustrative diagram of the force-distance relationship that may be employed in one arrangement of the method: electrostatic force vs. distance in

(A); direct contact hydrophobic force vs. distance in (B); and the conveyance of the hydrophobic force via equilibrium vapor bubble in (C).

Fig. 9 is an illustrative diagram of simulation steps that may be implemented in one arrangement of the method wherein a vapor bubble (910) is spontaneously generated in the vicinity of two hydrophobic residues (920) in a region known to form an alpha helix (A). The charged residue (930) can move away from the line-of-sight (a line charting the shortest distance between a reference point (e.g., the RMS center) of one hydrophobic residue to another hydrophobic residue) between the two hydrophobic residues by rotation allowing the vapor bubble to insert in the gap (B). The backbone moves in response to the forces generated by the coalescence (940) of the bubbles (C) forming an alpha helix loop segment (D). In (E) a beta sheet region vapor bubble insertion is blocked by one of the two similarly charged residues. Mutual repulsion prevents the movement of both charged residues to a region out of the line-of-sight between the two hydrophobic residues.

Fig. 10 is an illustrative diagram of the folding of Ubiquitin under different modeling conditions that may be implemented in one arrangement of the present invention: electrostatic interaction only in (A), hydrophobic forces represented by a single square well, plus electrostatic interactions in (B), bubble extended two square well hydrophobic forces in (C), bubble extended two square well hydrophobic forces with intervening charge conditional extension rule and electrostatic forces in (D) (1040 and 1060). A Pymol illustration of the experimentally determined Ubiquitin structure is provided for comparison in (E) (1050 and 1070).

Fig. 11 is an illustrative diagram comparing the simulation-generated protein structure against experimental NMR data for the same.

Fig. 12A is an illustrative diagram of simulation results using case D of Fig. 10 to determine the volume as a function of time in one embodiment of the invention.

Fig. 12B is an illustrative diagram of simulation results using case D of Fig. 10 to determine the total energy as a function of time in one embodiment of the invention.

Fig. 13 is a diagram of the most accepted structure of the nAChR M2 protein 1310, and an alternate nAChR M2 structure 1320 (commonly thought to exist at the onset of denaturizing) having an obvious kink 1330 in the central region of the alpha helix.

Fig. 14 is an illustrative diagram of the nAChR M2 protein structure generated using an embodiment of the invention incorporating a kinetic model wherein a positive two electron charge is placed near the central region of the alpha helix.

5 While the present invention will now be described in detail with reference to the Figs., it is done so in connection with the illustrative embodiments.

### DETAILED DESCRIPTION

The present invention provides for a method to determine the native structure of a polypeptide and/or protein, its structural response to an altered  
10 environment, and an altered new equilibrium structure resulting from the altered environment. The method comprises a protein model and a simulation algorithm involving one or more of thermodynamic energy type partitioning, an energy gradient, energy constraints and experimentally observable quantities including but not limited to viscosity, temperature, equilibrium vapor bubble concentrations, molecular  
15 diffusion and drift rates.

The invention may be used to predict polypeptide shape change and folding rates resulting from an environmental impetus including but not limited to protein-protein proximity and/or solution characteristics such as pH, temperature, illumination, electric and magnetic fields. For example, when external electric fields  
20 approach the (very large) values seen in the close proximity to a point charge (*e.g.*,  $1.4 \times 10^{11}$  N/C) the vapor filled alpha helix core of a protein may collapse if the energy stored in the polarized water can exceed energy gains realizable by coalescence of vapor/non-vapor interfaces. Magnetic permeability of the water can produce a similar effect in very strong magnetic fields. Light of an ionizing energy  
25 type can create new charged residues and/or species, thereby promoting and/or inhibiting alpha helix and/or beta sheet formation.

The analytic method in this invention is based on a kinetic description of molecular system evolution, which employs an equilibrium reference framework and a mode of motion along the free energy gradients, the speed of which is  
30 connected to the mobility of the moving particles. The route-mean-square or RMS volume (the average distance between the atomic constituents of the protein under

consideration), is used as a measurement of the amount of protein folding when starting from a known denatured linear state.

Consider the onset of folding in response to an environmental change. The protein begins in a known denatured state. When starting from a linear protein in ambient conditions consistent with a linear structure (for example a protein in liquid water near but below boiling temperature), said linear structure represents the structure with the lowest conformational free energy. The protein is now placed in ambient conditions consistent with a different, necessarily more compact structure than the previous maximum RMS volume linear state. Under the new environment, the old protein shape is no longer than of the lowest conformational free energy and therefore is unstable, and the protein will compact, thereby reducing its conformational free energy. The energy reduction is comprised of, but not limited to, electrostatic free energy, hydrophobic energies (dependent on the environment, not the protein structure), van der Waals forces, and entropy (such as the number of kink sites, vibration spectra, rotational and translational entropies). It is therefore convenient to consider said energy changes to be a function of the RMS volume during protein compacting or folding.

Protein folding and unfolding can be viewed as symmetrical responses of a protein to an environmental impetus. Even when such folding and unfolding is caused by rapid environmental change, it may be treated as a series of small environmental changes and corresponding symmetrical responses of the protein. The protein will take a path along the slope of free energy surface to achieve a new stable shape. If we have two or more distinct energy types, the conformation shape-energy surface for each type can be plotted, and the path constraint one energy-type partition imposes on the other is illustrated in Fig. 1, where the intersection of the two conformation shape-energy surfaces 110 and 120 define the available lowest energy state, and the path would be along the curve 130 that forms at said intersection, as illustrated in Fig. 1. Free energy gradients may be used to define the available lowest energy state, and the minimum and maximum found along the contour. Motion along the intersection curve involves changes (gradients) in both energy types. In this invention, said two distinct energy types can be intramolecular energies and background thermal energies.

The algorithm used in this invention is based on the cooperative motion of a large group of atoms about each of the Ramachandran (dihedral)  $\phi$  and  $\psi$  angles with a speed that is determined jointly by the mobility of the protein molecule, which mobility is dependent on the characteristics of the solvent, temperature and/or internal energy change with respect to the change in density of charged residues, and by one or more forces derived from internal energy force fields, including but not limited to bonding energies and hydrophobic interactions. Each dihedral angle can have independently a zero or non-zero bonding energy associated with rotation. In the algorithm, the bond rotation energy is typically set to zero. The combined forces acting on atoms, groups of atoms, and/or amino acid residues on opposing sides of the dihedral angle axis produce torques acting upon effective lever arms of the protein, thereby folding the protein to its native shape.

The present invention also provides for an algorithm implemented in a source code that translates said physical description into a series of commands that, upon input of a particular protein and/or polypeptide amino acid sequence; an optional input of temperature; an optional input of diffusion constant and/or an optional input of global energy change upon folding will produce a folded structure of said input amino acid sequence under the input parameters. Global energy change can be inferred from experimental results (as the energy released by the folding of a protein), or it may be treated as a parameter to be revealed by the simulation when examining a protein with a known structure or the known reaction of a protein structure to a known environmental impetus. Additionally, when both the structure and the global energy change are unknown, a number of potential structures predicated on different global energies may be generated. The span of these potential structures is in-turn predicated on the actual relative strength of the global energy to specific electrical, hydrophobic, or other defined energy terms.

The algorithm used in this method is based, in part, on physical drift and diffusion via change of the Ramachandran (dihedral) angles on the polypeptide or protein backbone, which corresponds to a cooperative (multi-atom) motion. As illustrated in Fig. 2, the algorithm starts from an arbitrary conformation and eventually transforms a protein molecule into its final folded state through successive steps which are linked to real time. In each step, the algorithm computes internal forces resulting from the particular conformation taken by the protein molecule as

well as an effective random force resulting from the random collision of the protein with the surrounding media, typically water. The algorithm then moves the molecular segments on either side of the alpha carbon bond pair having the largest net torque (sum of the actual force and the random effective thermal force multiplied by the appropriate lever arm length) relative to one another for the duration of a time step along a chosen dihedral angle by  $\theta$  ( $\theta = d/l$ , where  $d$  is the displacement of a moving particle in thermal diffusion, and  $l$  is the effective lever length of the protein).

A diagram of the calculation of torques on each backbone atomic pair is presented in Fig. 3. A cylindrical coordinated system is defined for each atomic pair, for example atomic pair  $j$ , and  $j+1$  of Fig. 3. Where the z-axis of said cylindrical coordinate system is defined in the direction of the bond between said atomic pair  $j$  and  $j+1$ .

In terms of said coordinate system the torque generated about said axis defined for said atomic pair  $j$  and  $j+1$  is a product of the torques and/or forces generated between every atomic pair  $i, j+n$  where  $i$  ranges from 1 to  $j-1$  and  $n$  ranges from 2 to the number  $N$  where  $M=N+j$  and  $M$  is the number of backbone atoms in the protein and/or polypeptide. That is, forces existent on atoms and residues on one side of the atomic pair  $j$  and  $j+1$  interact with those on the other side. Forces directly on atom (and residue if any)  $j$  and atom (and residue if any)  $j+1$  do not generate a rotational torque about said atomic pair axis and are therefore not counted in the force or torque generation.

Said forces and/or torque is predicated upon the charges attached to said pairs  $i, j+n$  found on either side of atomic pair  $j$  and  $j+1$  and/or the charges attached to residues attached to said pairs  $i, j+n$  and/or the degree of hydrophobicity of said residues attached to every said pairs  $i, j+n$ , other forces such as but not limited to hydrogen bond related forces may also be considered.

Torque generated about axis z for said atomic pair  $j$  and  $j+1$  is a product of the force generated by every said pair  $i, j+n$  for every allowed  $i$  and  $n$  value. Only the force components,  $F_{\theta_{i,j+n}}$ , in the direction  $\hat{\theta}$  produces torque about said axis z for said atomic pair  $j$  and  $j+1$ . Said forces may be dependent on said distance between said pairs  $i, j+n$ . The total torque about said axis may be defined for said atomic pair  $j$  and  $j+1$  as:

$$T_{j,j+1} = \sum_{i,j+n} F_{\theta i,j+n} \times \left( \frac{r_i + r_{j+n}}{2} \right)$$

$$F_{ave j,j+1} = \frac{T_{j,j+1}}{r_{ave}}$$

It is possible to accurately determine  $r_{ave}$  by summing the values of  $r_i$  and then dividing by the number  $j-1$  to determine a value of  $r_{iave}$  and then summing the values of  $r_{j+n}$  and then dividing by  $M-J-1$  to establish a value for  $r_{j+nave}$  followed  
 5 by the averaging of two said average values. The various average length,  $r_{ave}$  may be set as half of the length of the denatured protein or the approximate radius of a partially compacted protein in order to speed calculation. Other fixed or variable  $r_{ave}$  values such a number that is average for proteins of a given type or size are also may also be used. Other reasonable schemes for calculation of the torque and the  
 10 radius arm through which it is applied can be used.

The simulation moves by rotation about one axis  $z$  for one said atomic pair  $j$  and  $j+1$  per time step. The particular axis  $z$  for said particular atomic pair  $j$  and  $j+1$  selected for rotation on a given time step is determined by tabulating the values of  $T_{i,i+n}$  for every backbone pair and then adding a random torque value consistent with  
 15 thermal energy,  $kT$ , the product of the Boltzmann constant and the Kelvin temperature used for the simulation.

In one time step there are  $M-1$  axis associated with  $M-1$  torque calculations to each a random thermal energy is added yielding an adjusted torque values,  $T_{i,i+n}$ , that are then tabulated. The atom pair having the largest adjusted torque  
 20 value is allowed to move for a time period,  $t$ , according to said angular displacement formula:

$$\theta \approx \sin \theta \approx \frac{\mu F_{ave i,i+n} t}{r_{ave}}$$

Where  $\mu$  is the mobility applying to force type  $F_{ave j,j+1}$  which may be a function of a second energy consideration and other terms such as but not limited to  
 25 the viscosity of the environment as discussed below.

This mode of move on a protein molecule is illustrated in Fig. 4 (wherein amino acid side chains are represented as aromatic rings). These hydrophobic residues (410 and 420) come into direct contact, whereupon the two

bubbles or internal surfaces of the water/hydrophobic interface coalesce, thereby exerting a force (as indicated by lines 430) on a distant alpha carbon pair under consideration (440). The torque about the bond axis of alpha carbon pair 440 generated by the coalescence of hydrophobic residues 410 and 420, as well as any  
 5 reduction of that torque caused by the coalescence, and all other possible alpha carbon bonds that may be influenced by the coalescence is calculated as in Fig. 3. The global energy also involves density dependent entropy. In cases where this entropy cannot be determined from experimental results, a trial move is applied, the approximate density calculated using an average mobility, and the change in entropy determined,  
 10 then the actual rotation is performed using an appropriately modified mobility. The process is repeated until a fixed number of time steps have been performed or the structure ceases to change conformation with respect to predetermined volume criteria (for example, a fraction of the RMS volume) for a fixed number of time steps.

In the above procedure, the displacement  $d$  of a moving particle in  
 15 thermal diffusion is a function of the diffusion constant and time  $t$ ,  $d \approx \sqrt{Dt}$ , where  $D$  is the diffusion constant and  $t$  is the time step used in the simulation, e.g.,  $10^{-7}$  s, which when sufficiently short invokes the preceding approximations. A reasonable range for  $t$  is from  $10^{-13}$  s to  $10^{-3}$  s. The diffusion constant used may be the known diffusion constant of a similarly sized protein. Alternatively, the diffusion coefficient  
 20 could be treated as a simulation fitting parameter that can be used to adjust the amount of folding that occurs in a particular time step. Also alternatively, the diffusion coefficient could be calculated from first principals such as the Einstein approach (e.g.,  $D_{rms} = \sqrt{\frac{kT}{3\pi\eta a}} \sqrt{t}$  where  $\eta$  is the viscosity of the liquid,  $a$  the radius  
 of the protein,  $k$  is Boltzmann constant,  $T$  the temperature in Kelvin, and  $t$  is time. See  
 25 Taylor, Zafiratos, Dubson, Modern Physics for Scientists and Engineers 2<sup>nd</sup> ed. Prentice Hall Inc NY 2004, p.105).

The effective random force such that the same displacement would be produced as if the protein backbone atoms were undergoing random thermal diffusion is then determined as a Gaussian distribution about an average effective force

30  $F_{eff} \approx \pm kT \sqrt{\frac{1}{Dt}}$  (because  $d = \mu F_{eff} t$ ), where  $k$  is Boltzmann's constant and  $T$  is the

temperature in Kelvin degrees. Correspondingly, the average torque (the product of

force and distance) that acts upon about each Ramachandran axis

is  $F_{eff} \times r = \pm r_0 kT \sqrt{\frac{1}{Dt}}$  where  $r_0$  is the average length of the lever arm through which

the force acts. In the case of Ubiquitin a value of  $10^{-6} cm$  is reasonable. When  $D$  is  $10^{-5} cm^2/s$ ,  $t$ , one computational time step, is  $10^{-7} s$ , and  $T$  is  $300^{\circ}k$  a torque value of

5  $\frac{kT}{cm}$ , is obtained.

Another approach yielding a similar result is based upon the force necessary to exactly counteract diffusion. Since drift and diffusion must be balanced at equilibrium,  $D\nabla n = \mu F_{eff} n$ , where  $n$  for this consideration is defined as  $n = \frac{N}{V_m}$  and

where  $N$  is the number of atoms in the protein and  $V_m$  is the volume of the molecule

10 with respect to the Ramachandran angle as seen in Fig 5,  $V_m = \int_0^{\theta} \int_0^{\delta} \int_0^r r dz d\theta dr = \frac{r^2}{2} \theta \delta$ .

Here  $\theta$  is the Ramachandran angle under consideration,  $r$  is the average length of the protein as measured from the axis of rotation and  $\delta$  is the average thickness of the wedge shaped volume, as illustrated in Fig. 5. Therefore, volume is predicated on angle change since

$$n = \frac{N}{\left(\frac{r^2}{2} \theta \delta\right)}$$

15

$\therefore$

$$\nabla n \approx \frac{1}{r} \frac{d}{d\theta} \left( \frac{N}{\frac{r^2}{2} \theta \delta} \right) = \frac{-N}{\left(\frac{r^3}{2} \theta^2 \delta\right)}$$

Substitution of this equation into the balance of drift and diffusion also using the Einstein relation yields:

$$D \frac{-N}{\left(\frac{r^3}{2} \theta^2 \delta\right)} = \mu \left( \frac{N}{\frac{r^2}{2} \theta \delta} \right) F_{eff}$$

$\therefore$

$$F_{eff} = \frac{-D}{\mu \theta} = -\frac{kT}{r \theta}$$

Thus the torque acting on the axis of the Ramachandran angle corresponding to this effective force is  $F_{eff}r = kT/\theta$ , where  $\theta$  runs from 0 to  $\pi$  with a central value of  $\pi/2$  being a reasonable estimate producing a torque value approximately equal to  $kT$ .

5 Energy-Coupled Mobility Determination

Einstein relations express the strict relationships between field driven drift mobility and diffusion constant for probabilistic particle motion (*See* Albert Einstein, *Annalen der Physik*, 17, 549 (1905); Richard J. Borg and G.J. Dienes, “*An Introduction to Solid State Diffusion*” Academic Press, NY 181 (1988) (“Borg & Dienes”); Paul G. Shewmon, *Diffusion in Solids*, McGraw-Hill, NY 24 & 140 (1963) (“Shewmon”). For example, the energy conservation consideration can be applied to the steady-state distribution of mobile electrons under the attraction of an opposite charge at a fixed position. The steady-state diffusion current (electrons diffusing away from the attractor) balances the drift current (electrons moving towards the

10 attractor), leading to a net zero energy transport. Here an external electric field is present. In another example, the charges carried by an atomic site on a protein molecule may move via energy absorbing (or releasing) reconfiguration. A charged protein region moving relative to the rest of the protein produces a changed structure characterized by changed distributions of kinks, twists and/or rotated bonding sites.

20 Consider the simple case of a system having a mobile charged species, with no matrix coupling, and moving under the influence of an electric field (the gradient of an electric potential). Using Onsager relation (e.g., see Borg & Dienes and Shewmon), the drift current  $\vec{J}_{drift}$  of an ensemble of charged particles having a density  $n_p$  and in an electric field  $\vec{E}$ , is:

25 
$$\vec{J}_{drift} = -n_p \mu \vec{E} \quad (1)$$

where  $\mu$  is the mobility of the charged species with respect to the electric field. Fick’s relation for the random-walk-like diffusion current,  $\vec{J}_{diff}$ , of this charged species is proportional to the product of the gradient in species density (or concentration),  $\nabla n_p$  and a diffusion constant  $D$ :

$$\vec{J}_{diff.} = -D\nabla n_p \quad (2)$$

The total current,  $\vec{J}_{Total}$  at a particular location is

$$\vec{J}_{total} = -n_p \mu \vec{E} - D\nabla n_p. \quad (3)$$

5 In steady state, the total flux must be zero:

$$\vec{J}_{total} = 0 = -n_p \mu \vec{E} - D\nabla n_p. \quad (4)$$

10 Therefore:

$$n_p \mu \vec{E} = -D\nabla n_p \quad (5)$$

or

15

$$\mu \vec{E} = -D\nabla (\ln(n_p)). \quad (6)$$

20 Next, the electric field is defined as the gradient of the electrostatic potential energy:

$$\vec{E} = \nabla \phi_{pot.} \quad (7)$$

Combining equations 6 and 7 yields:

$$\bar{E} = \nabla \phi_{pot.} = -\frac{D}{\mu} \nabla (\ln(n_p)). \quad (8)$$

Defining an electrostatic chemical potential in terms of the density of charged particles,  $n$ , and the electrostatic potential,  $\phi_{pot.}$ :

$$\phi_{echem.} = \phi_{pot.} + kT (\ln(n_p)). \quad (9)$$

5                   The first term on the right hand side of equation 9 is the internal energy term (e.g., the Fermi level for electrons and holes) and the second term is the entropy of mixing term. Equilibrium requires that both the net current fluxes and the chemical potential gradients of the relevant charged species be zero whenever there is no compensating gradient in the matrix chemical potential:

$$10 \quad \nabla \phi_{echem.} = \nabla \phi_{pot.} + kT \nabla (\ln(n_p)) = -\frac{D}{\mu} \nabla (\ln(n_p)) + kT \nabla (\ln(n_p)) \quad (10)$$

Equation 10 is solved in the case of a zero matrix free energy change  $\nabla \phi_{echem}=0$ :

$$\frac{D}{\mu} = kT \quad (11)$$

15                   This is the well-known Einstein relation. When there is an energy coupling between charged species and the matrix in which in moves, the matrix free energy consideration must be formulated. Following the definition of thermodynamic terms used by Dickerson, the free energy,  $G$ , and the change in free energy,  $dG$ , induced by a change in the environment (i.e., external electric fields, temperature, pressure, etc.) can be expressed in terms of chemical potentials as:

$$20 \quad G = -ST + VP + \sum_i \phi_{echem_i} n_i + \sum_j \phi_{L_j} n_j + \sum_k \phi_{O_k} n_k \quad (12)$$

and

$$dG = -SdT + VdP + d \left\{ \sum_i \phi_{echem_i} n_i \right\} + \sum_j \phi_{L_j} dn_j \quad (13)$$

where in equation 13 the contribution by charged particles to the change in free energy is expressed in the summation of  $\sum \phi_{echem}$ , and the chemical potentials related to non-electrostatic effects such as strain fields in the material matrix are expressed in the second summation,  $\sum \phi_L$ , whereas all other chemical potentials related to the matrix (but not a function of charge particle density in any way) are represented in the final summation. (See Richard E. Dickerson, *Molecular Thermodynamics*, W.A. Benjamin, Inc., Menlo Park, CA 249 (1969) (“Dickerson”).

While the atmospheric pressure dependence of  $VdP$  is not expected to be a large factor in solid matrix systems, the induced strain fields may be. Therefore, the present formulation retains the term  $VdP$ . Note that the full differential of the electrostatic chemical potential must be maintained because changes in the local and global electric field affect the free energy contribution of these charged species and the electric field will be an implicit consideration of the ensuing formulations.

With an eye towards spatial dependent perturbation applications (such as a system in an applied external field) the free energy gradient generated between a reference state (and/or location within the material matrix) and a perturbed state (and/or location within the material matrix) is determined:

$$\nabla G = \nabla \left\{ VP - ST + \left( \sum_i \phi_{echem_i} n_i + \sum_j \phi_{L_j} n_j \right) \right\} \quad (14)$$

Equation 14 provides a mathematical basis for a system in equilibrium and is assumed to be applicable to a system perturbed slightly from equilibrium. In equilibrium the free energy has the same minimum value at any arbitrary position in the system, so

$$\nabla G = \nabla \left\{ VP - ST + \left( \sum_i \phi_{echem_i} n_i + \sum_j \phi_{L_j} n_j \right) \right\}_{\forall x_i, y_i, z_i} = 0 \quad (15)$$

Since every point in a protein (or polypeptide) is in thermal contact with every other point, and since the protein is free to change shape (to a degree free to move), when motion ceases or begins to fluctuate about a given shape, it is necessarily the case that forces have ceased. That is, the forces (generally free energy gradients) arising from a steady-state and/or equilibrium state sum to zero. Therefore a small localized (non-

uniform) change in the environment (e.g., a localized temperature change) that induces a change in the energy stored in the matrix strain fields about charges results in motion of charged particles. These motions cease when an equilibrium condition is established that has a net zero free energy gradient as well as having a net zero  
 5 current. Interestingly, the protein's shape may oscillate about a given shape, but the energy of this oscillation must in turn be consistent with the temperatures ( $\sim kT$ ). Such oscillation is physical and possible in real protein systems and is possible in the simulation.

The speed of a return to steady-state (equilibrium here) when a  
 10 perturbation is introduced can be used to define mobility. The speed of a charged particle moving in a free energy gradient (and/or an electric field) is dependent on both the free particle speed and the speed at which the matrix can respond to charged particle redistribution. When in true equilibrium, this responding speed can be obtained by assuming it is equal to the speed of motion of the charged particles.  
 15 Using equation 10 and equation 15 yields:

$$\nabla G = \left\{ -\frac{D}{\mu} + kT \right\} \nabla \ln n_p + \nabla \left\{ VP - ST + \left( \sum_j \phi_{L_j} n_j \right) \right\}_{\nabla x_i, y_i, z_i} = 0 \quad (16)$$

Equation 16 expresses the fundamental relationships between kinetic factors that equilibrate fluxes of charge particles and thermodynamic factors that couple the matrix to the particles. Rearranging equation 16 and solving for the  
 20 mobility yields:

$$\mu = \frac{D \nabla \ln(n_p)}{\nabla \left\{ VP - ST + \sum_j \phi_{L_j} n_j \right\}_{x_1, x_2, x_3 \dots x_n} + kT \nabla \ln n_p} \quad (17)$$

When there are no strain or other couplings to the matrix (i.e., when the value of the terms bracketed in the denominator), equation 17 reverts to the well-known Einstein relation,  $\mu = D/kT$ .

25 Use of this energy-coupled mobility by a simulation of Ubiquitin, a single chain polypeptide composed of 74 amino acids, is demonstrated here as a non-limiting example. The simulation is conducted according to the algorithm described

above, and the mobility is determined according to equation 17 to account for global changes induced by shape change. In a hypothetical situation in which only entropy changes with volume, where  $\nabla H = 0$ , and where  $\nabla S$  can be expressed as  $n_k$  (here the product of the number of kink sites and the entropy value for a site):

$$5 \quad \nabla S = \nabla n_k \approx \frac{1}{r} \frac{\partial n_k}{\partial r} \quad (18)$$

Simplification of equation 17 yields

$$\nabla \ln n_p \approx \frac{1}{r} \frac{\partial \ln n_p}{\partial r} \approx \frac{1}{r} \frac{\partial \ln n_p}{\partial n_k} \frac{\partial n_k}{\partial r} \quad (19)$$

Therefore equation 17 becomes:

$$\mu \approx \frac{D}{kT \left( 1 - \frac{\partial n_k}{k \partial \ln n_p} \right)} \quad (20)$$

10 Thus obtained mobility  $\mu$  is incorporated into the algorithm described above.

Fig. 6 shows a linear illustration of Ubiquitin in (A); the result of a folding simulation using a constant mobility  $\mu = D/kT$  in (B), where diffusion constant  $D$  is obtained from Dwight E. Gray, Coordinating Ed., *American Institute of Physics Handbook 3rd.*, McGraw Hill, NY, 2-206 (1972); and the result of a folding

15 simulation using a variable mobility  $\mu$  (in equation 20) where  $\frac{\partial n_k}{k \partial \ln n_p}$  is assigned the values of  $-0.6$  under contraction and  $+0.6$  under expansion (in order to illustrate the impact of this parameter on structure generation), in (C). The corresponding faster contraction relative to expansion results in a more compact structure with a volume of  $0.33 \pm 0.16$  (arbitrary units averaged over 20 simulations) compared to  $0.54 \pm 0.23$  for

20 the constant mobility case. The speed of motion is important because it impacts the distance moved in a given time step, the motion always choosing the largest force according to the simulation design. The largest speed produces the largest displacements, which in-turn produce the greatest energy reduction. Therefore higher

speed in one direction relative to another will produce a structure automatically having a lower energy.

The energy term in the mobility equation (here the entropy term) is also considered. The entropy term attempts to express an energy reduction (which the formulation accounts for in speed) as a result of a more compacted structure (where there is greater entropy and therefore lower free energy) when steady-state is reached. It is possible to determine the free energy of folding (minus the electrostatic energy change) entirely or in part by experiment, such as from measurement of the heat of folding minus an inferred electrostatic energy change predicated on the net electrostatic energy difference between folded and denatured states. The free energy of folding may also be dropped from the energy term if the energy term is dominated by an entropy term that increases upon compaction. It is possible to use the simulation to deduce an effective free energy change (minus the electrostatic energy) by trial and error simulation of protein folding, by matching folded structures and/or folding rates for a protein for which sufficient experimental data is available. Note that neither structure shown here correlates well to the known structure because this example was focused solely on the use of global energy terms. More sophisticated simulations incorporating more force considerations including hydrophobic interactions between protein residues are necessary to produce more accurate molecular shape.

#### Hydrophobic Interactions Considerations

The well-understood hydrophobic interaction results from water surface produced at interfaces with non-polar and/or un-charged media. Owing to the highly polar nature of the water molecule, such surfaces extract significant energy cost. Each hydrophobic residue on the protein molecule has an effective bubble around it.

Consideration of equilibrium phenomena, such as the vacancy density in solid-state diffusion (see e.g., Shewmon) is typical of the kinetic description. Here the necessary equilibrium consideration involves the density of background bubbles, because the bubbles provide a mechanism for force conveyance. When two bubbles come in intimate contact a force is generated by the surface tension which tends to reduce the surface area of the two bubbles through coalescence.

Fig. 7 shows the probability of the natural occurrence of vapor bubbles in water as a function of bubble radius ( $2 \times$  distance). The probability is an equilibrium volume probability and it is assumed that there is a bubble of a given radius in a particular water volume when the probability is large (e.g.,  $> 0.5$ ), in a manner similar to the vacancies that are present in metals. See Richard J. Borg and G.J. Dienes, *An Introduction to Solid State Diffusion*, Academic Press, NY. When contacting the effective bubbles that encase each of the two hydrophobic residues (Fig. 8) the natural bubble provides a conduit to convey force. Paralleling the formulation of the vacancy in solids described by Hull (David Hull, *Introduction to Dislocations 2nd Ed.* Pergamon Press, NY, 16 (1975)), the equilibrium probability of bubbles (shown in Fig. 7) of a given radius  $V_B$  in water can be calculated:

$$\frac{V_B}{V_x} = K = e^{-\Delta G/kT} \quad (21)$$

where  $\Delta G$  is the free energy change for equilibrium between bubbles of size  $V_B$  and  $V_x$ , which can be expressed as

$$\Delta G = \int_0^{r_0} P dV + \int_0^{r_0} \frac{2\gamma}{R} dS \quad (22)$$

where  $PdV$  is the work needed to expand a bubble filled with an equilibrium water pressure,  $V_x$  is the equilibrium probability of the same volume being filled with water,  $\gamma$  is the surface tension of water ( $\sim 72$  dynes/cm),  $r_0$  is the radius of the bubble under consideration, and  $dS$  is an elemental surface. See Dwight E. Gray, Coordinating Ed., *American Institute of Physics Handbook 3rd.*, McGraw Hill, NY, 2-206 (1972). For a spherical bubble, the elemental surface can be written as:  $dS = \frac{8}{3} \pi R dR$ . Bubble densities can also be experimentally measured and/or inferred as a function of bubble radius from direct measurement and analysis of the water or solution under consideration. Also, it is possible to consider bubbles of different shapes and/or the expansion of a given bubble to generate an appropriate force versus distance map for the interaction of two hydrophobic residues. The simulation may also be used to construct an effective force versus distance force map for the interaction of two hydrophobic residues by trial and error simulation of protein folding, through

matching folded structures and/or folding rates for a protein for which sufficient experimental data is available.

Bubble formation is affected significantly by electric fields produced by charges, since the electric field tends to draw in the strongly polar dielectric water thereby lowering the energy stored in the electric field. The corresponding energy  
 5 necessary to remove the water molecules polarized by the electric field must be paid during bubble formation, and is dependent upon the shape of the charged residue or specie, the distance about the charges the water needs to be removed from, the  
 10 magnitude of the charge, and the distance at which the charged residue or species can be approached by water.

The strength of the electric field created by the charge is dependent upon the distance to the location under consideration,  $r$ , and the relative polarizability

$$\vec{E} = \frac{k_c q}{\chi r} \hat{r}$$

of the media  $\chi$  (usually water, where  $\chi = 74$ ): , where  $q$  is the charge under consideration in Coulombs (C),  $r$  is the distance between the charge and the  
 15 location under consideration, and  $\hat{r}$  is the direction in a convenient coordinate system between said charge and said location expressed as a unit vector, and  $k_c$  is the Coulomb constant.  $\chi$  will vary depending upon the distance,  $r$ , as the presence of any water between the charge and the location under consideration is not certain. At 0.1  
 20 nm, it is not physically possible for there to be any water in between the charge and the location, so  $\chi = 1$ . At distances of less than  $\sim 1$  nm, it is possible to calculate the electrical energy reducing polarization by calculating the probabilistic presence of a water molecule and the degree to which it is polarized, so  $\chi$  is between 1 and 74. At large distances,  $\chi$  can be assumed to be 74.

Accordingly, equation 21 must be modified to reflect the extra energy  
 25 needed to remove the dielectric (water) from the electrostatic field:

$$\frac{V_B}{V_X} = K e^{-W/kT} \quad (23)$$

where  $W$  is the work required to transfer a bubble of size  $V_x$  from the electric field in water to an electric field in vapor (essentially vacuum). (John D. Jackson, *Classical Electrodynamics, 3rd edition*, Wiley & Sons, NY, 167 (1999)).

Under the condition where a charged residue is characterized by a single electronic charge having a minimum radius of 2 angstroms (diameters smaller than the size of a water molecule are meaningless) and a maximum radius of 1.0 nm (the distance at which bubbles without electrostatic fields begin to show markedly  
5 decreased probability as seen in Fig. 7),  $W$  is approximately 2.5 eV. Correspondingly the product  $Ke^{-W/kT}$  is essentially zero as shown in Fig. 8. Under such condition, bubbles become improbable and hydrophobic force cannot be conveyed in regions having electric fields in the line-of-sight between two hydrophobic residues.

The hydrophobic force extension in the presence of an electric field is  
10 represented as double wells dependent on distance (one-half of the radius) between two charged particles, as shown in Fig. 8. Obstruction of the line-of-sight occurs when the mutual repulsion of several similarly charged residues prevent the rotation of all of the charged residues away from the line-of-sight region as shown in Fig. 9. If the space between two neighboring hydrophobic residues is sufficiently close such  
15 that the effective bubble and/or water/non-polar interface encasing each will wither directly touch and/or come into mutual contact with an equilibrium vapor bubble, the touching will generate a force that pulls the two residues closer together. This occurs via a set of rotations of the alpha carbon bonds between the two residues.

In the event a hydrophilic (polar) residue or other charged specie is  
20 attached to the protein in a position between the two residues, the energy stored in the polarized water surrounding it will seek to remain about the residue or charge specie. Accordingly, the residue or specie will experience a force tending to rotate it away from a vapor region or the hydrophobic residues encased in effective bubbles. This rotation can be inhibited by factors such as steric hindrance or electric fields created  
25 by other charged species.

Take, for example, a situation wherein two or more similarly charged residues interfere with each other as they attempt to rotate away from a vapor-filled region, thereby remaining surrounded by polarized water (or other media). Due to their proximity, when the residues rotate to the same side of the protein backbone  
30 chain, the distance between them will decrease. This distance change corresponds to a significant energy increase due to Coulomb repulsion. It is reasonable to expect that this energy is sufficiently large so as to overcome energy reductions made possible by

hydrophobic residues coming close enough to reduce the water/non-polar and/or vapor interface.

It should be noted that two hydrophilic residues of opposite or nearly opposite charge will automatically rotate to the same side of the backbone chain so as to reduce the distance between them. The presence of two such residues between two hydrophobic residues would not inhibit extension of the non-polar/vapor interface necessary for the generation of hydrophobic force since the two can rotate as a unit. The electric field produced by two hydrophilic residues diminishes much faster than that of a single point charge. In some ideal cases it may be that two hydrophilic residues of opposite or nearly opposite charge have a sufficiently small electric field (and therefore small amount of surrounding polarized water/media) that the surrounding polarized water is stripped away, allowing the two hydrophilic residues to occupy a space within a vapor-filled region directly between two neighboring hydrophobic residues.

Computationally, the obstruction condition for the blocking the interaction of two neighboring hydrophobic residues by an number of hydrophilic and/or charged species is satisfied when

$$\left| \sum_i q_i \right| = \sum_i |q_i| \quad (25)$$

where  $q_i$  is the charge on  $i$ -th charged residues located in the region between the two hydrophobic residues. Under such condition, the intervening charged residues have the same sign of charges and therefore would be inhibited from moving to one side of the backbone chain to allow the formation of an effective bubble, such that the long-range hydrophobic force  $h_2$  between the two residues, shown in Fig. 8, is inapplicable.

Physical consideration of the bubble extension of the hydrophobic force is a function of the actual residues involved, whether or not said residues are hydrophobically bound to other hydrophobic residues (thereby said double or triple hydrophobic residues can extend further for a given thermal energy,  $kT$ ), and a function of environment (such as but not limited to temperature, ionic densities, pH, presence of surfactants, and the presence of other proteins, polypeptides, molecules, or atoms. Said force map representation of said bubble extension of the hydrophobic force can have multiple extensions predicated by said considerations.

Again we choose Ubiquitin to illustrate the effect of hydrophobic interactions in protein folding simulation. The initial state was a linear representation of the Ubiquitin backbone where the charge and degrees of hydrophobicity of each of the side chain amino acid residues are known (e.g., by using methods such as the  
5 comparison of heats of solution in polar and non-polar solvents, see Robert A. Copeland, *Methods for Protein Analysis, A Practical Guide to Laboratory Protocols*, Chapman & Hall, NY, 14 (1994)), and their values were assigned to appropriate locations on the protein backbone chain.

Fig. 10 shows the results of Ubiquitin simulations under various  
10 hydrophobic force parameters using a constant mobility. Case (A) shows the simulation result 1010 using only electrostatic driving forces. The simulation did not produce an alpha helix region. Case (B) shows is the simulation result 1020 using both electrostatic forces and hydrophobic forces, where only a short-range square well hydrophobic force in Fig. 8 is used. The protein structure obtained has unnatural and  
15 large loops that do not easily dissociate. Extending the operating range of hydrophobic force to 0.75 nm with a second shallower square well in Fig. 8 in case (C) enhances the probability that two nearly spaced hydrophobic residues experience a hydrophobic force sufficient to result in motion under the conditions of the simulation in the helix region as the second and third neighbor hydrophobic residues  
20 react strongly and early in the simulation prior to loop formation, which is evidenced by the observation in 1030 of a central counter loop 1035 formed by an alpha helix. Note, there may exist hydrophilic residues between said hydrophobic residues. However, unnatural hairpin loops form in a beta sheet region under this condition. Depicted at 1040 and 1060 is the result obtained in case (D) using the conditional  
25 line-of-sight charged residue consideration of equation 25. It is observed that beta sheet regions remain flat and alpha helix regions form loops with orientation consistent with a smooth unidirectional helix, which is consistent with experimental results obtained from NMR imaging at 1050 and 1070.

The average simulation time for producing the secondary structure is  
30 of the order of  $10^{-5}$  s to  $10^{-4}$  s and consumes about ten minutes of CPU time on a standard desktop computer. Note that the simulation time needed is dependent on the diffusion constant used, and more accurate determination of the diffusion constant would lead to a better approximation of the folding time. The calculated folding times are within the 8 ms time frame beyond which regions of Ubiquitin become

inaccessible to the proton exchange mechanisms Briggs and Roder used to characterize folding times. *See* Martha S. Briggs and Heinrich Roder, *Proc. Nat. Acad. Sci.* 89, 2017-2023 (1992). Also the result 1040 case (D) in Fig. 10 is in agreement with the collapse characteristics described by Jacob *et al.* Jaby Jacob, 5 Bryan Kranz, Robin S. Dothager, P. Thiyagarajan, and Tobin R. Sosnick, *J. Mol. Bio.* 338, 369-382 (2004). A further comparison of the simulation result to experimental data obtained from NMR imaging is shown in Fig. 11

Figures 12A and 12B show the time evolution of the RMS (root-mean-square) energy and volume, respectively, for the case where the both the 10 electrostatic and conditional hydrophobic extension forces were applied. The energy in this case was calculated by assuming an arbitrary initial energy and subtracting the work done,  $F \cdot \theta \cdot r_0$ , done in each time step where  $F$  is the force applied (the sum of hydrophobic, electrostatic, and thermal forces),  $\theta$  is the calculated angle change, and  $r_0$  is the average lever arm length of the protein.

15 At the onset of folding there are large forces along the line-of-sight between various charged locations and large hydrophobic forces particularly in the regions that will develop an alpha helix. However, the resolved forces in the direction of Ramachandran angle rotation are small. The nearly spaced residue hydrophobic interactions soon dominate the energy selection algorithm steps. The volume change 20 and the energy released in the alpha helix formation are small as these involve a relatively small fraction of the protein atoms and because the distance over which the forces act are short. Because of the shape of the hydrophobic force map (Fig. 8) it is necessary to provide a halt after a fixed number of steps (typically ten time steps) and freeze the relative position of the hydrophobic near neighbor residues for a fixed time 25 period so that the algorithm will proceed to sample through the next largest forces. After the initial incubation time the protein quickly collapses as the Ramachandran rotation axis come into favorable orientation with respect to the large electrostatic and hydrophobic torques.

30 At long times thermal fluctuations again dominate with small volume and energy changes. These will not damp out in a simulation involving real temperature and are considered to be representative of a real protein in a liquid water environment at room temperature. Nonetheless, the inclusion of the hydrogen

bonding and other considerations consistent with tertiary structure would be expected to reduce the volume excursion but not necessarily the energy excursions.

Referring to Fig. 13, the nAChR M2  $\alpha$ -helix region was modeled in order to demonstrate the ability of one embodiment of the invention to generate secondary structure. The M2 is one of a large number of alpha helix regions that line the ion channel in nAChR. Although it is known that the basic structure of M2 is an  $\alpha$ -helix 1310, there is some debate as to whether a kink or super-coiled region 1330 of the alpha helix exists 1320 in the central region during the gating of ionic flow. See Law Richard J., Forrest, Lucy R., Ranatunga, Kishani M., Rocca, Paola La, Tieleman, Peter, and Sansom, Mart S. P., *Proteins: Structure, Function, and Genetics*, 39, 47-55 (2000).

The algorithm was applied to the nAChR M2 structure. Under nominal residue charge states, an un-kinked structure 1410 similar to that of Fig. 13 was generated as shown in Fig. 14. However, if one or two charges are added to the central region, a kink 1430 appears as shown in 1420. The driving force for the kink formation can be easily understood in terms of the forgoing discussion of the model and its physical basis. Namely, the presence of charge draws water a highly polar fluid with a correspondingly large dielectric constant, into the vapor filled alpha helix wherever the energy reduction presented by water filling a dielectric region is greater than the energy increase related to increased vapor-water and non-polar residue-water interface.

#### Environmental Effects

The simulation may be used to track the possible effects of various environmental changes. Environmental changes can have profound effects on protein secondary structure. For instance, illumination can, when the photon energy is sufficient, excite an electron of an atom in the protein and/or polypeptide and/or within water and/or other environmental media and/or species thereby changing the capacity of said atom with said excited electron to react with other elements or compounds within water, the protein and/or polypeptide. These reactions include the direct bonding of one element to another. This direct bonding can lead to local (to the bonding site(s)) changes in charge, hydrophobic character, and/or but not limited to flexibility of local bonds. Such changes can be tracked by the simulation.

Higher energy light (perhaps but not limited to photons with energy greater than  $\sim 2$  eV) could lead to direct ionization of elements within water, materials within the environment, and/or the protein and/or polypeptide. This ionization leads directly to a change in the charged locations within the protein and/or polypeptide.

- 5 The changes in-turn lead to changes in electromotive forces present in folding and folded proteins. Such changes can result in altered structure. Also ionization and the resultant charge change could lead to an alteration in local hydrophobic character thereby induces changes in hydrophobic interaction and structure.

The native structure of a polypeptide and/or protein in response to pH  
10 can be predicted using the simulation. pH can effect the concentration of ions in water and/or the media in which said polypeptide and/or protein under consideration is embedded. For example, pH can affect the presence of ions that turn effect the vapor/media (and/or water) and the hydrophobic residue/media (and/or water) interface. The interfacial energy in-turn has an effect on said surface tension of said  
15 interface. Some of the affects of pH are an increase in surface tension (promoting coalescence of bubbles and therefore increased hydrophobic force) while increased surface tension reduces the radius of equilibrium bubbles and increasing the energy needed to expand an existing bubble and/or envelop about a hydrophobic residue. This increase in energy corresponds to decreased probability of a given bubble  
20 displacement as well as a corresponding decrease in the distance at which the hydrophobic force can operate (to attract two hydrophobic residues toward one another). Also, pH can affect the probability of a solution hydrogen ion (and/or atom) attaching and/or bonding to the polypeptide and/or protein structure. Hydrogen bonding in-turn affects the charge structure of said polypeptide and/or protein thereby  
25 introducing a change that can be tracked by the simulation.

Importantly hydrogen ions and/or atoms can chemically bond to one location of the polypeptide and/or protein structure and then to a second location, thereby providing a new motive force drawing two said bonding locations together. Hydrogen bonding in-turn affects the charge structure of said polypeptide and/or  
30 protein thereby introducing a structural change that could be tracked by a modified version of the simulation.

The native structure of a polypeptide and/or protein in response to salts can also be predicted. Salts can effect the concentration of ions in water and/or the media in which said polypeptide and/or protein under consideration is embedded. For

example, salt ions can in-turn effect the vapor/media (and/or water) and the hydrophobic residue/media (and/or water) interface. This interfacial energy in-turn has an effect on said surface tension of said interface, which can have the previously described effects.

5                   The effects of charged species (pH related ions and/or salt related ions) can include a reduction of long-range interaction of protein charges on one another through the charge screening.

                  The native structure of a polypeptide and/or protein in response to non-polar, hydrophobic chaperons and/or small proteins existent in the environment and/or solution can also be predicted. Non-polar, hydrophobic chaperons and/or small  
10                   proteins existent in the environment and/or solution can effect the concentration of effective bubbles present in the environment in which said polypeptide and/or protein under consideration is embedded. Since each small chaperons and/or small proteins are encased in an effective bubble the environment will be significantly changed. For  
15                   example, one effect can be increased vapor/water interface area and vapor volume resulting in altered local electric field due to a change in the environmental dielectric constant and/or polarization. A change in the dielectric parameter would tend to increase the force arising from charge-charge interaction.

                  The increase in effective bubble density that includes bubble-encased  
20                   non-polar, hydrophobic chaperons and/or small proteins may increase the range of the hydrophobic force through mechanisms described for natural equilibrium vapor bubbles. At very high concentrations of non-polar-hydrophobic chaperons and/or small proteins the force between neighboring hydrophobic residues on a particular protein and/or polypeptide under consideration may be reduced as a high  
25                   concentration of chaperons and/or small proteins may displace sufficient amounts of water and/or other media so as to alter the surface energy throughout the local environment. Under conditions of reduced surface energy hydrophobic force and interaction is reduced. In extreme cases some structures stabilized and/or held together by hydrophobic interaction could denature and/or change structure.

30                   Similarly, the native structure of a polypeptide and/or protein in response to polar, hydrophilic, and/or charged chaperons and/or small proteins existent in the environment and/or solution can be predicted. Polar, hydrophilic, and/or charged chaperons and/or small proteins existent in the environment and/or solution can affect local electric fields and/or nearby concentrations of effective

bubbles and/or water (and/or other media) – hydrophobic (non-polar) residue interface in water (and/or the media) in which said polypeptide and/or protein under consideration is embedded since each such small chaperons and/or small protein induces a local electric field similar to that of an ion and/or point charge. At very high concentrations said chaperons and/or small proteins could markedly alter local protein structure and/or denature said protein structure through screening of interaction between charge-charge interaction and by decreasing the probability of vapor bubbles where ever the field of said chaperons and/or small proteins is sufficiently strong.

10                   The native structure of a polypeptide and/or protein in response to metals can also be predicted. Metals can act as a concentration of ions in the water and/or the media in which said polypeptide and/or protein under consideration is embedded, producing effects similar to that described for ions and salts above. Also, metals (especially heavy metals) have electron clouds extending beyond 0.1nm. A far-extending electron cloud can lead to charge screening resulting in modified charge-charge interaction within the protein and/or polypeptide under consideration. Metals and other elements can bond directly to the protein modifying local charge status and in-turn local hydrophobic character. These changes in local charge status can lead to altered charge-charge interaction and hydrophobic residue-hydrophobic residue interaction.

20                   The effects on protein structure induced by species such as surfactants known to modify interfacial and/or water/vapor surface energies can be predicted using the simulation. The effect of mutations or viruses on protein and/or polypeptide structure can also be predicted using the simulation when and where they induce a change in local charge status, and/or local hydrophobic character, and/or the distance between neighboring hydrophobic residue pairs, and/or the screening charge-charge interaction, and/or the distance between charges, etc.

25                   The foregoing merely illustrates the principles of the invention. Various modifications and alterations to the described embodiments will be apparent to those skilled in the art in view of the teachings herein. It will thus be appreciated that those skilled in the art will be able to devise numerous techniques which, although not explicitly described herein, embody the principles of the invention and are thus within the spirit and scope of the invention.

We claim:

1. A method for determining an equilibrium structure of a protein in a predetermined environment, the protein having Ramachandran angles and a known denatured structure, comprising:
  - (a) determining a maximum RMS volume of the known denatured structure of the protein;
  - (b) calculating at least one force on the protein in its current structure in the predetermined environment;
  - (c) determining a net torque resulting from the at least one force for each of the Ramachandran angles of the protein;
  - (d) rotating at least one section of the protein structure on a side of a Ramachandran angle with greatest torque as determined in (c) to form a new structure;
  - (e) determining a new RMS volume for the new structure; and
  - (f) repeating (b) to (e) until the new RMS volume of the new protein structure at (e) is not less than the RMS volume of the structure of the protein at (b).
2. The method of claim 1 wherein (b) to (e) are repeated until a fixed number of iterations have elapsed.
3. The method of claim 1 wherein the at least one force comprises the electrostatic force.
4. The method of claim 3 wherein the at least one force further comprises thermal noise.
5. The method of claim 3 wherein the at least one force further comprises the hydrophobic force.
6. The method of claim 5 wherein determining the hydrophobic force includes determining a probability of formation of natural bubbles.

7. The method of claim 5 wherein the at least one force further comprises thermal noise.

8. The method of claim 1 wherein the equilibrium structure of the protein in a predetermined environment is known (d) comprises:

(i) determining a global energy change upon folding of the protein; and  
(ii) rotating at least one section of the protein structure on a side of the Ramachandran angle with the greatest torque as determined in (c) to form a new protein structure.

10

9. The method of claim 8 further comprising:

(iii) calculating an approximate density using an average mobility; and  
(iv) determining a change in entropy from the approximate density.

15

10. The method of claim 9 further comprising:

(v) modifying the mobility in (i) based on the change in entropy of (iv).

11. The method of claim 10 further comprising:

(vi) repeating the rotation of (ii) using the modified mobility of (v).

20

12. A synthetic protein, the synthetic protein having a predetermined equilibrium structure in a predetermined environment, wherein an amino acid composition of the synthetic protein is determined using the method of claim 1.

25

13. The synthetic protein of claim 12, wherein said synthetic protein configured to move an ionic species against a concentration gradient.

14. The synthetic protein of claim 13, wherein the ionic species is sodium, and the concentration gradient is that encountered during the desalination of water.

30

15. The synthetic protein of claim 13, wherein a structure of the synthetic protein is based upon to a naturally occurring neural channel structure.

16. The synthetic protein of claim 13, wherein the synthetic protein may be stacked to move the ionic species against a larger concentration gradient.
17. A synthetic protein, the synthetic protein having a predetermined structure change in response to a predetermined environmental change, wherein an amino acid composition of the synthetic protein is determined using the method of claim 1.
18. A method of determining a global energy change upon folding of a protein having Ramachandran angles and a known denatured structure and a known equilibrium structure, comprising:
- (a) determining a maximum RMS volume of the known denatured structure of the protein;
  - (b) calculating at least one force on the protein in its current structure in the predetermined environment;
  - (c) determining a net torque resulting from the at least one force for each of the Ramachandran angles of the protein;
  - (d) rotating at least one section of the protein structure on a side of a Ramachandran angle with greatest torque as determined in (c) to form a new structure;
  - (e) determining a new RMS volume for the new structure;
  - (f) repeating (b) to (e) using different values for the global energy change upon folding until the new protein structure at (e) approximates the known equilibrium structure.

25

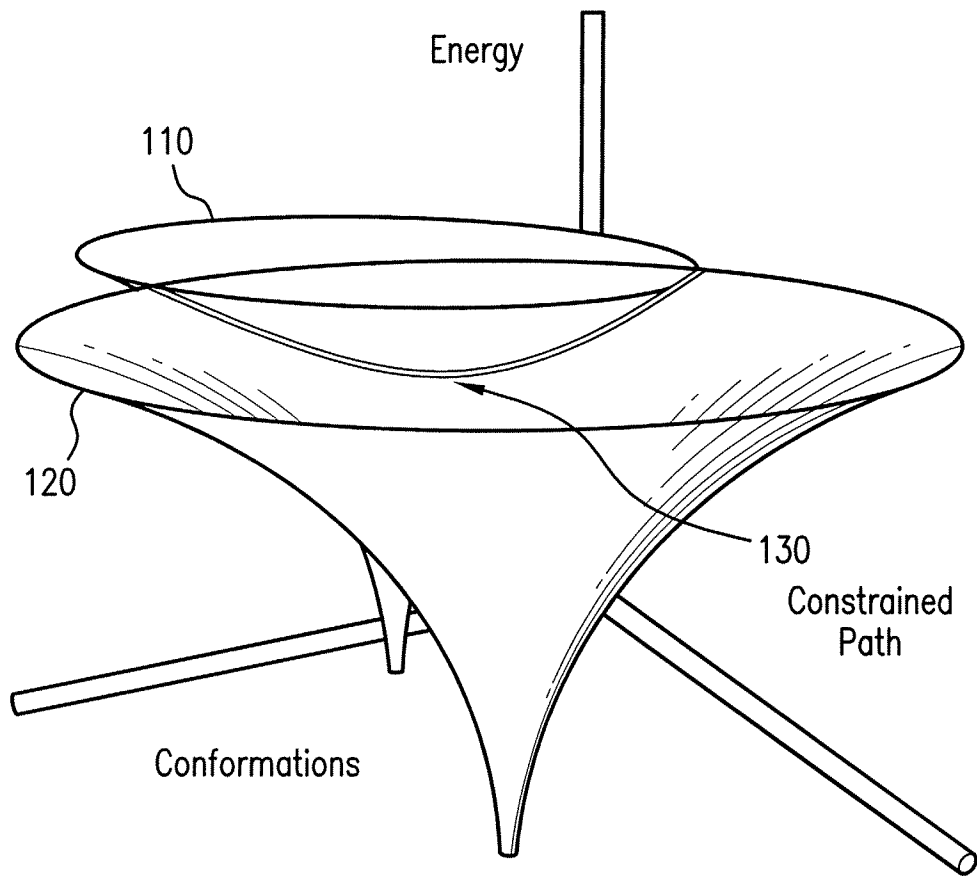


FIG. 1

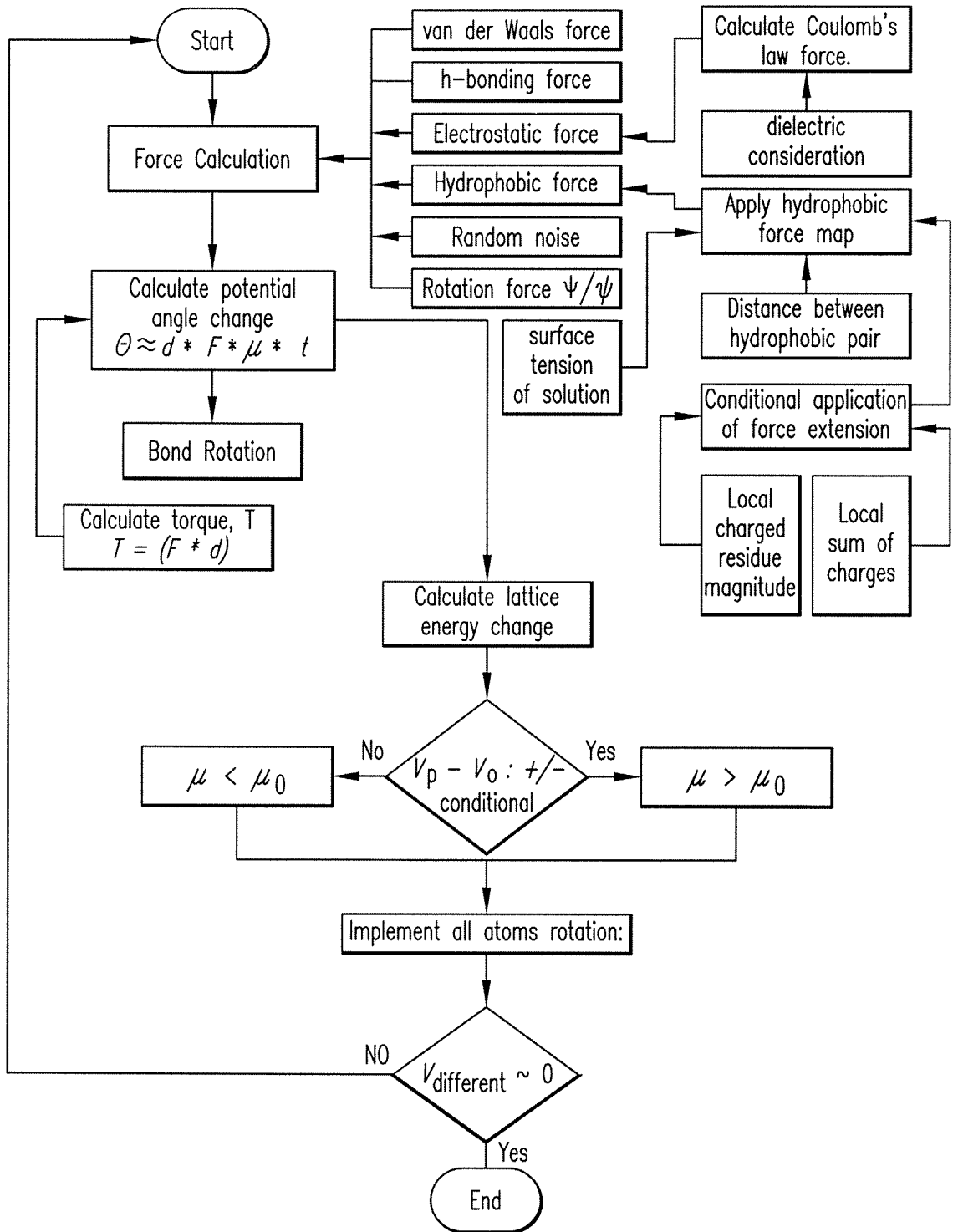


FIG.2

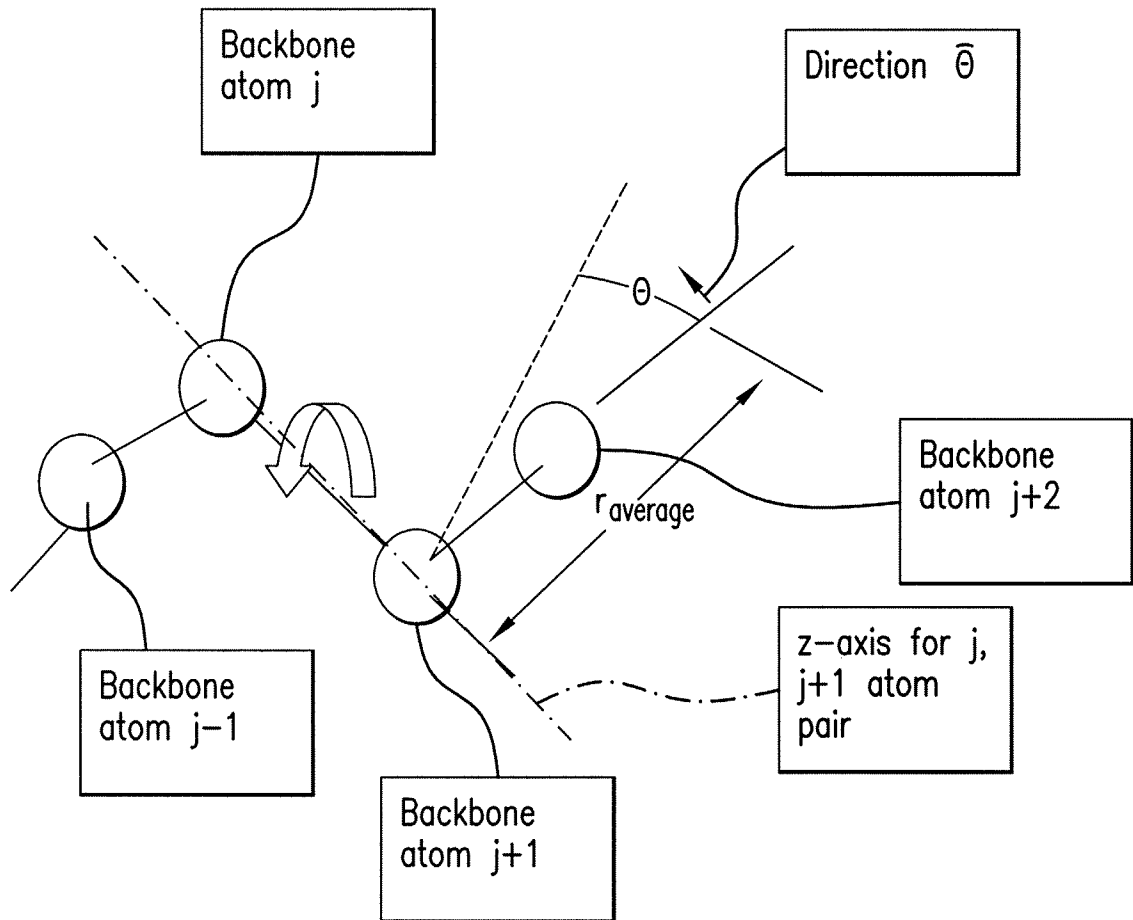


FIG.3

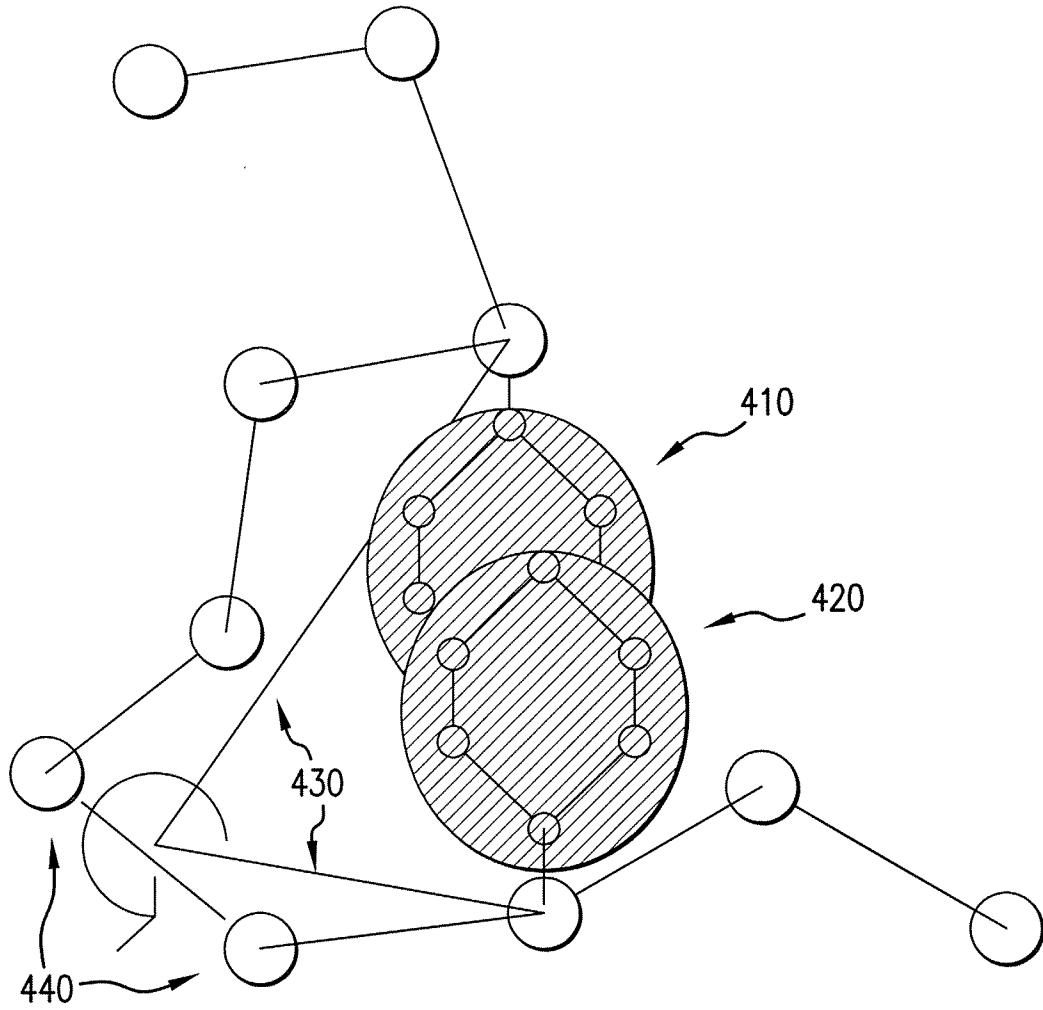


FIG.4

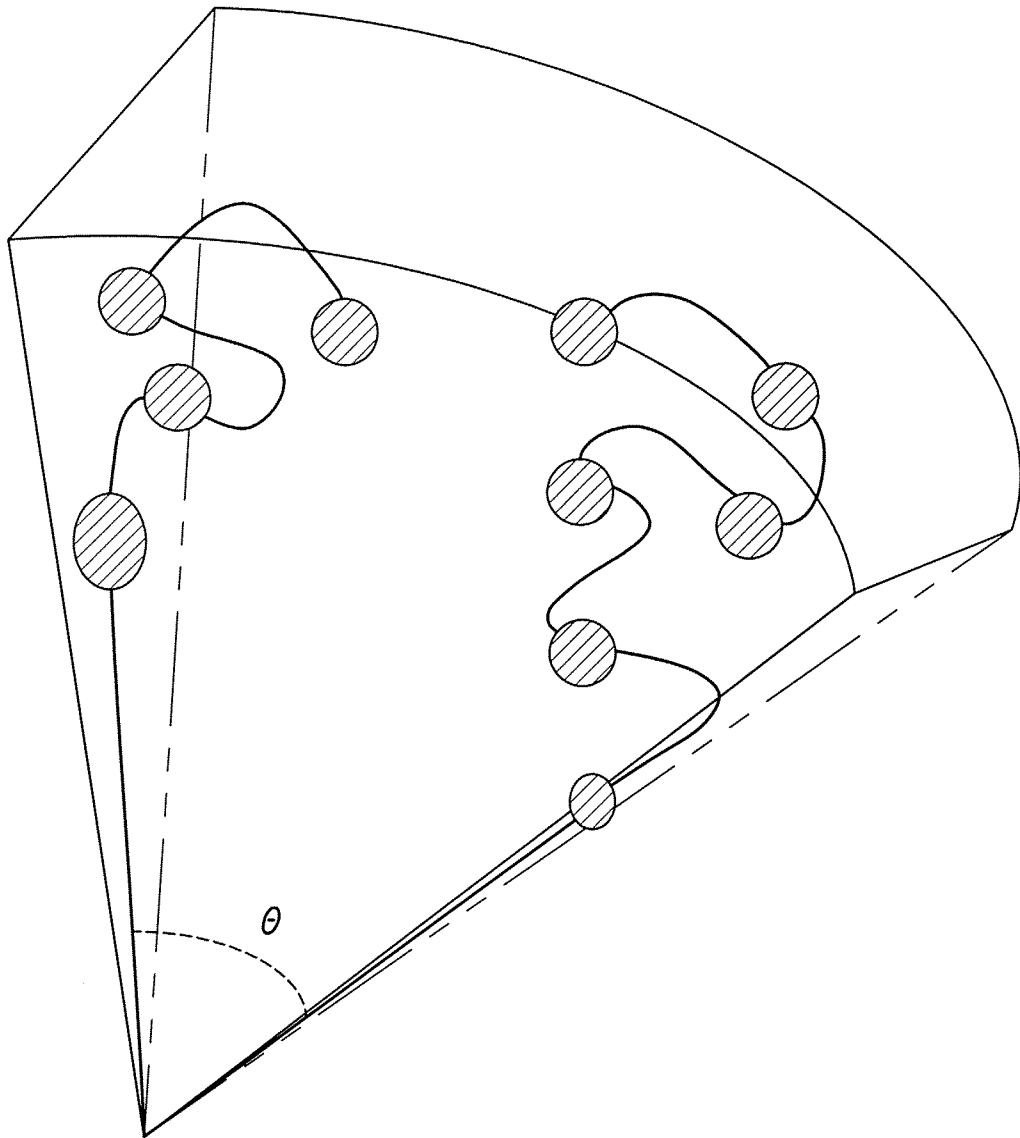


FIG. 5

6/16

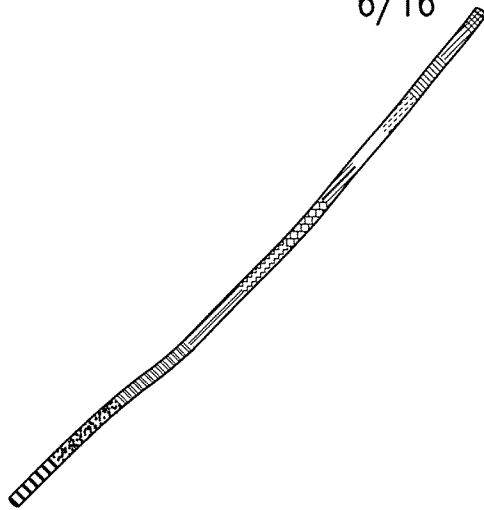


FIG. 6A

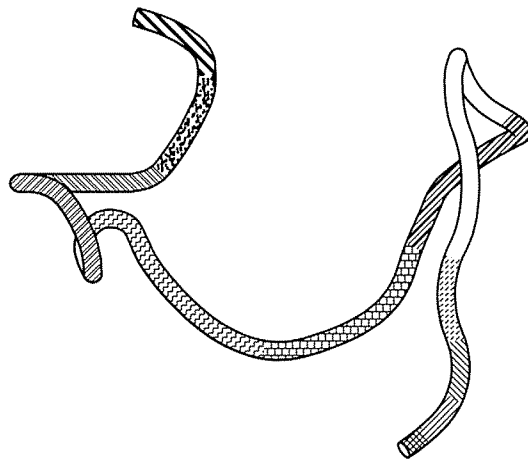


FIG. 6B

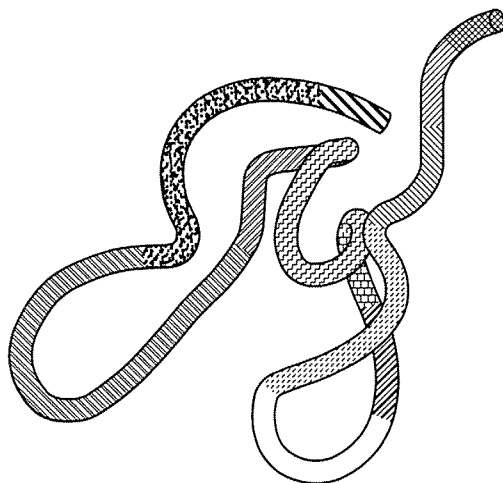














FIG. 6C

-  Dark Red
-  Red
-  Orange
-  Yellow-Orange
-  Yellow
-  Yellow-Green
-  Light Green
-  Green
-  Blue-Green
-  Light Blue
-  Blue
-  Dark Blue

7/16

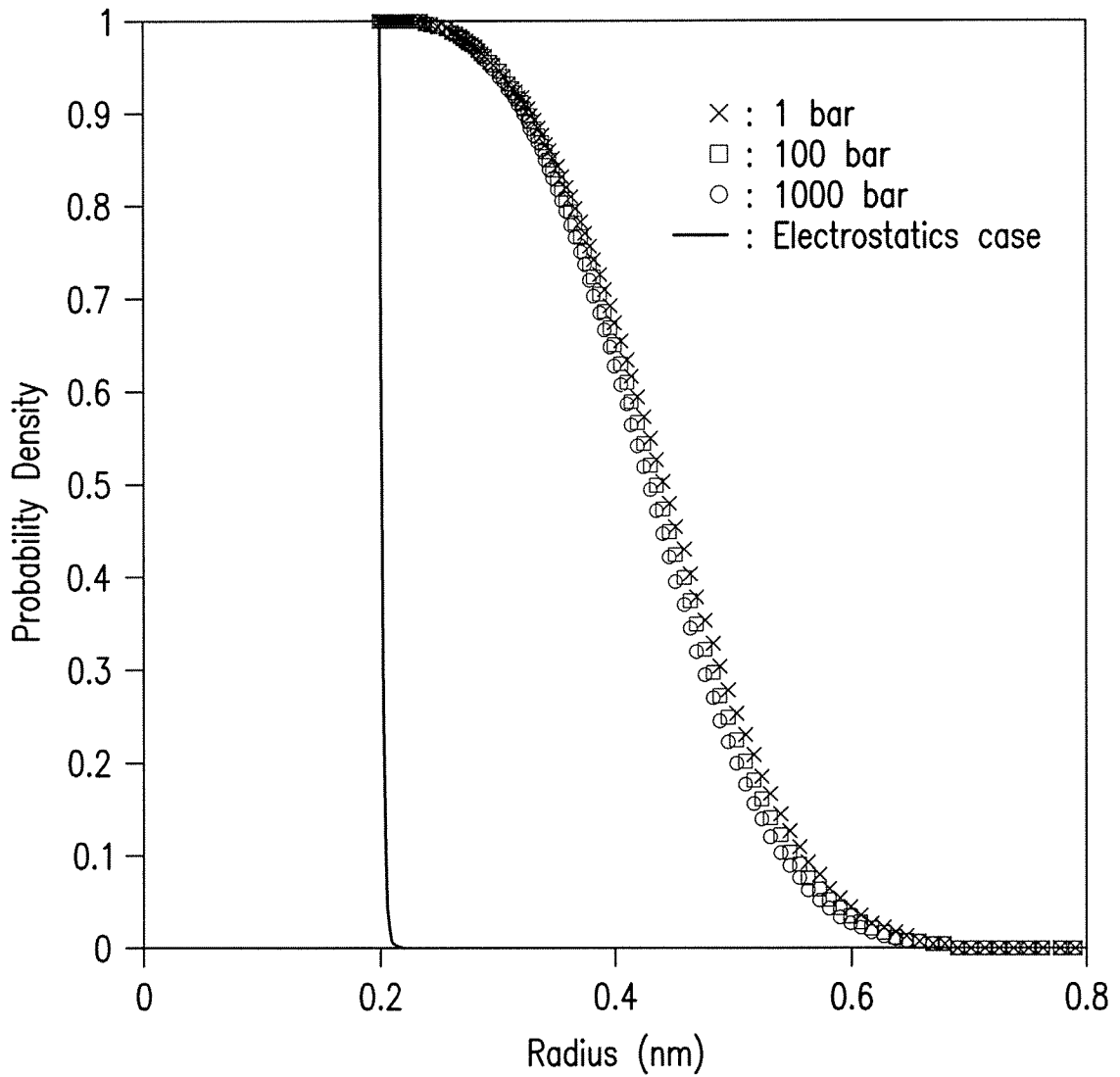


FIG. 7

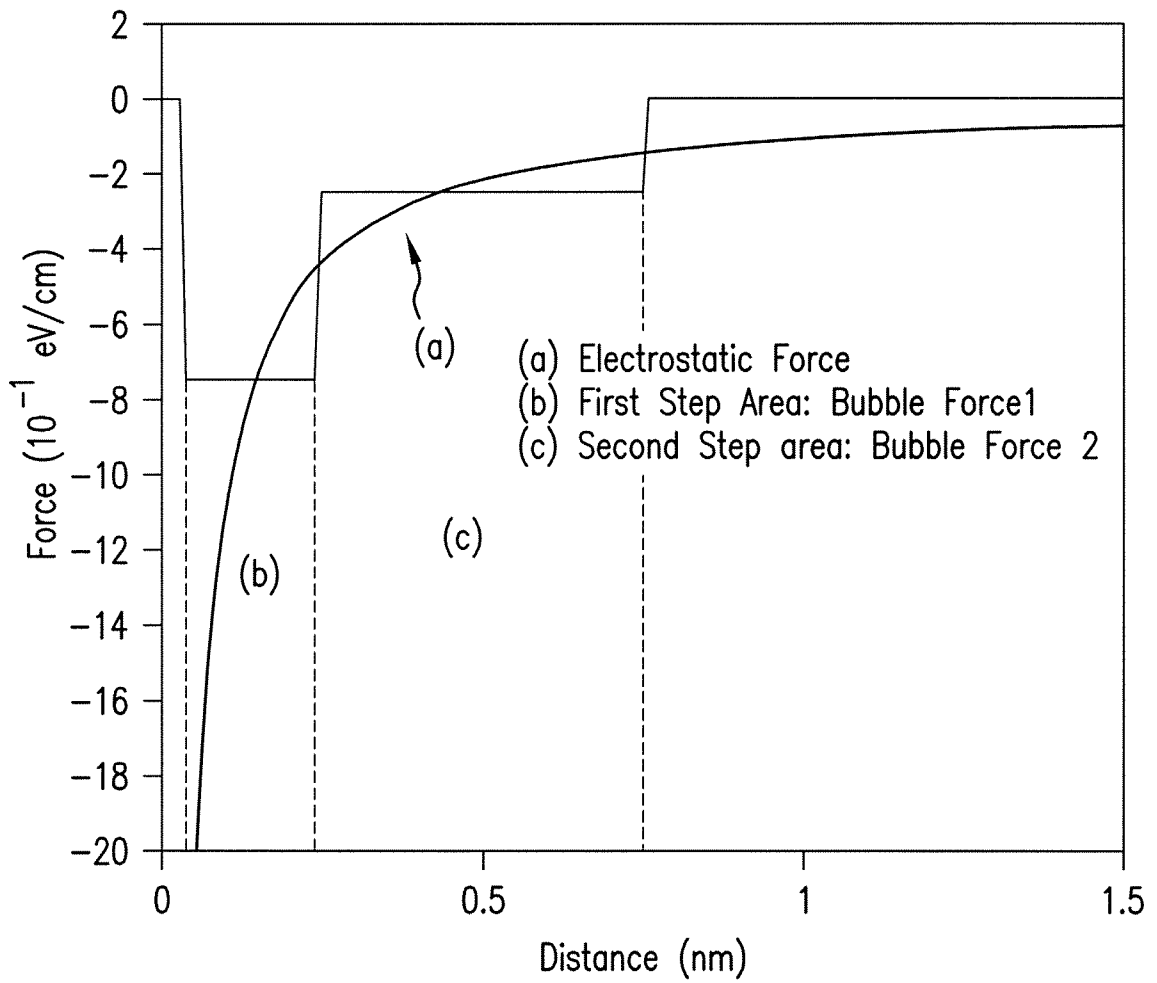
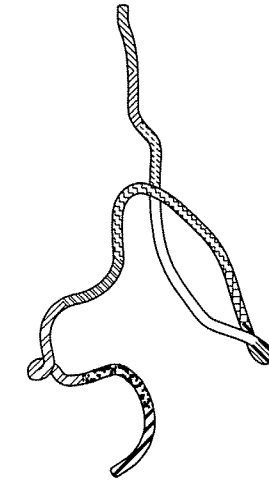
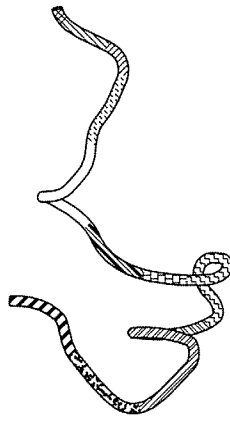
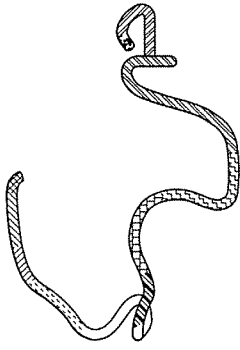
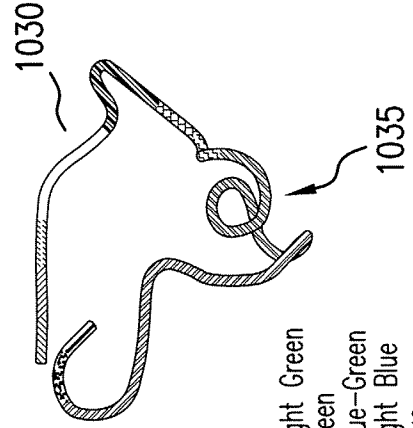
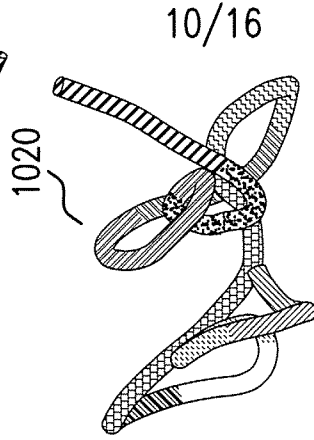
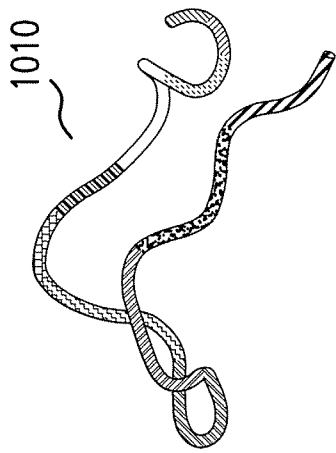


FIG. 8



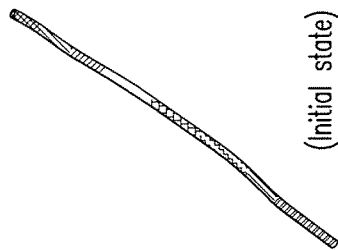


	Dark Red		Light Green
	Red		Green
	Orange		Blue-Green
	Yellow-Orange		Light Blue
	Yellow		Blue
	Yellow-Green		Dark Blue

FIG. 10A

FIG. 10B

FIG. 10C



(Initial state)

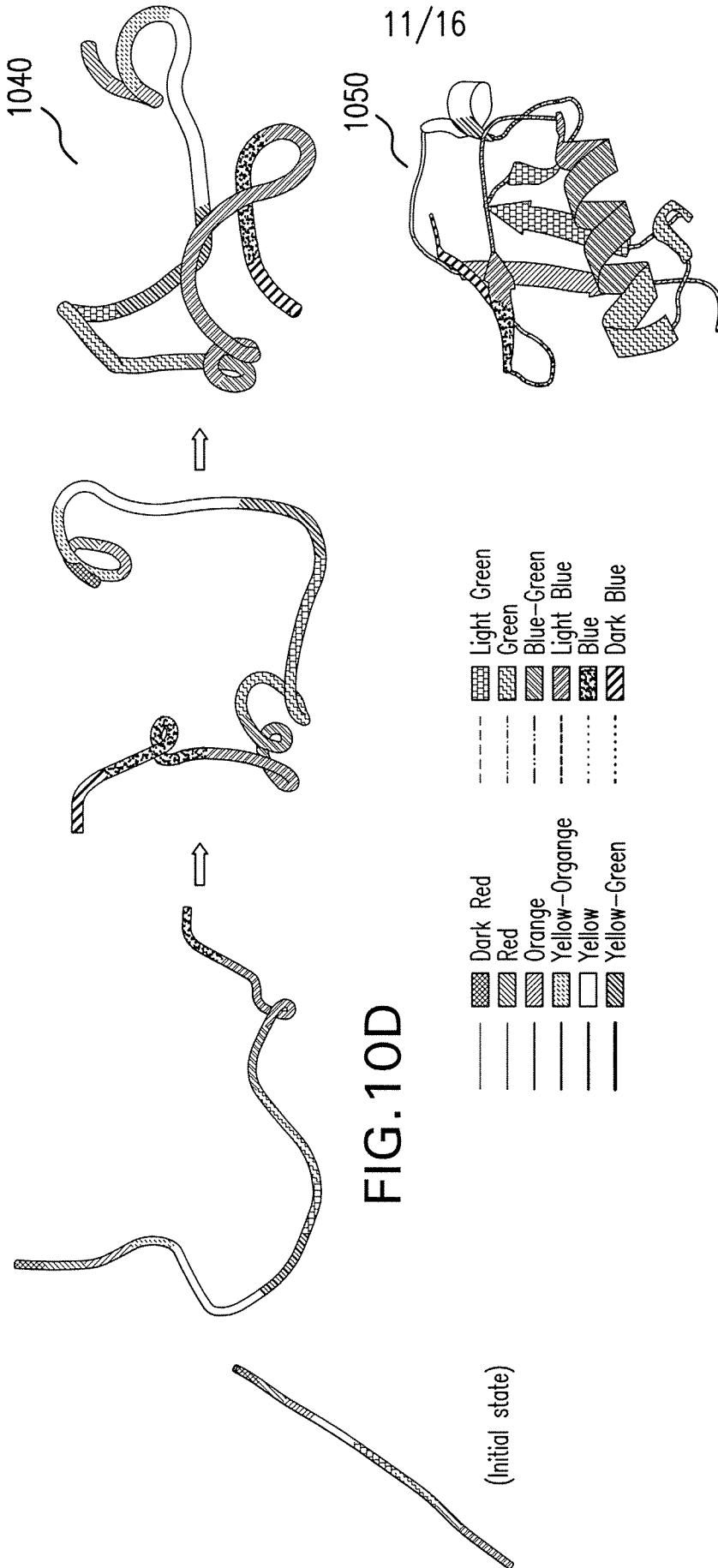


FIG.10D

FIG.10E

12/16

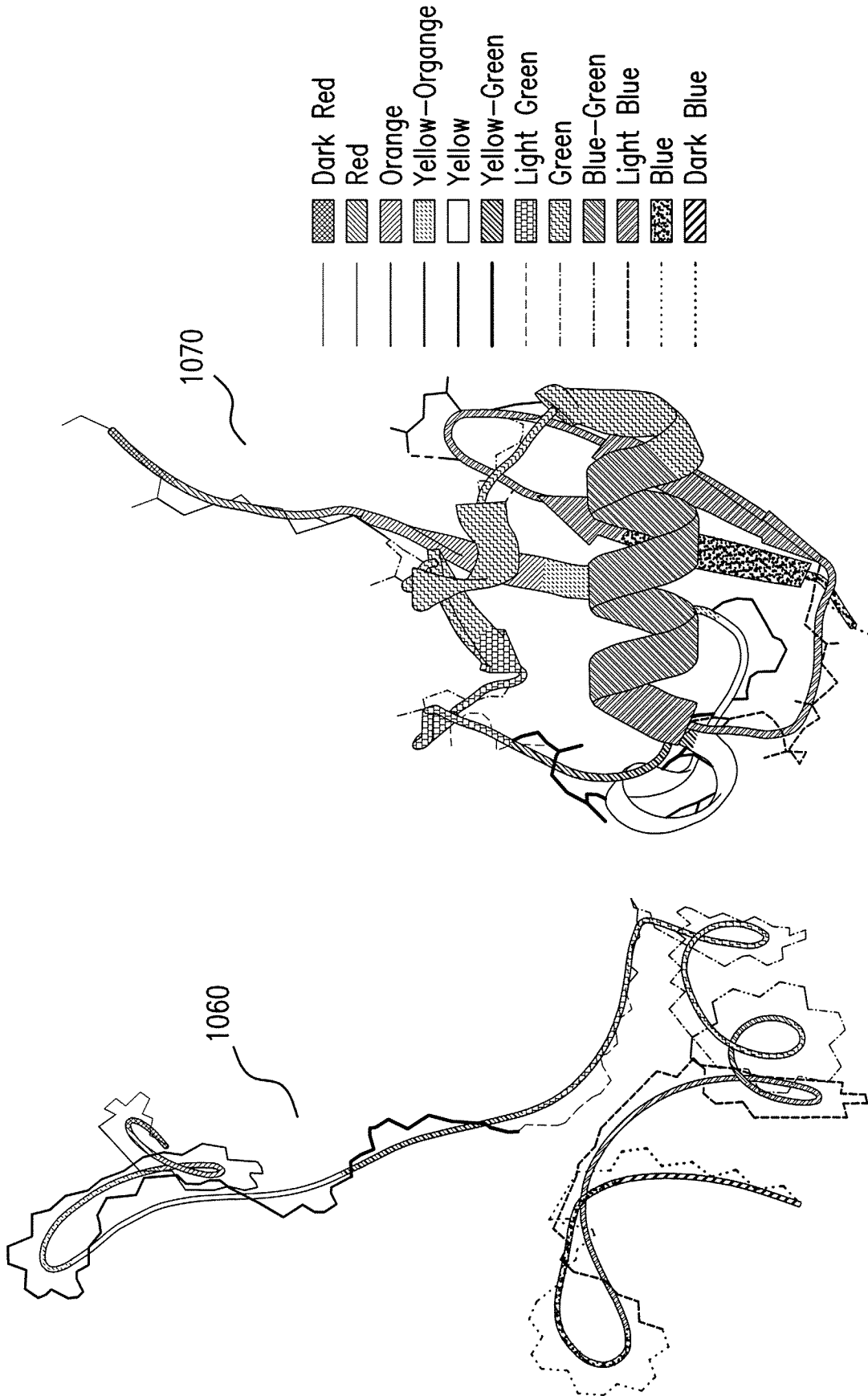


FIG. 10F

13/16

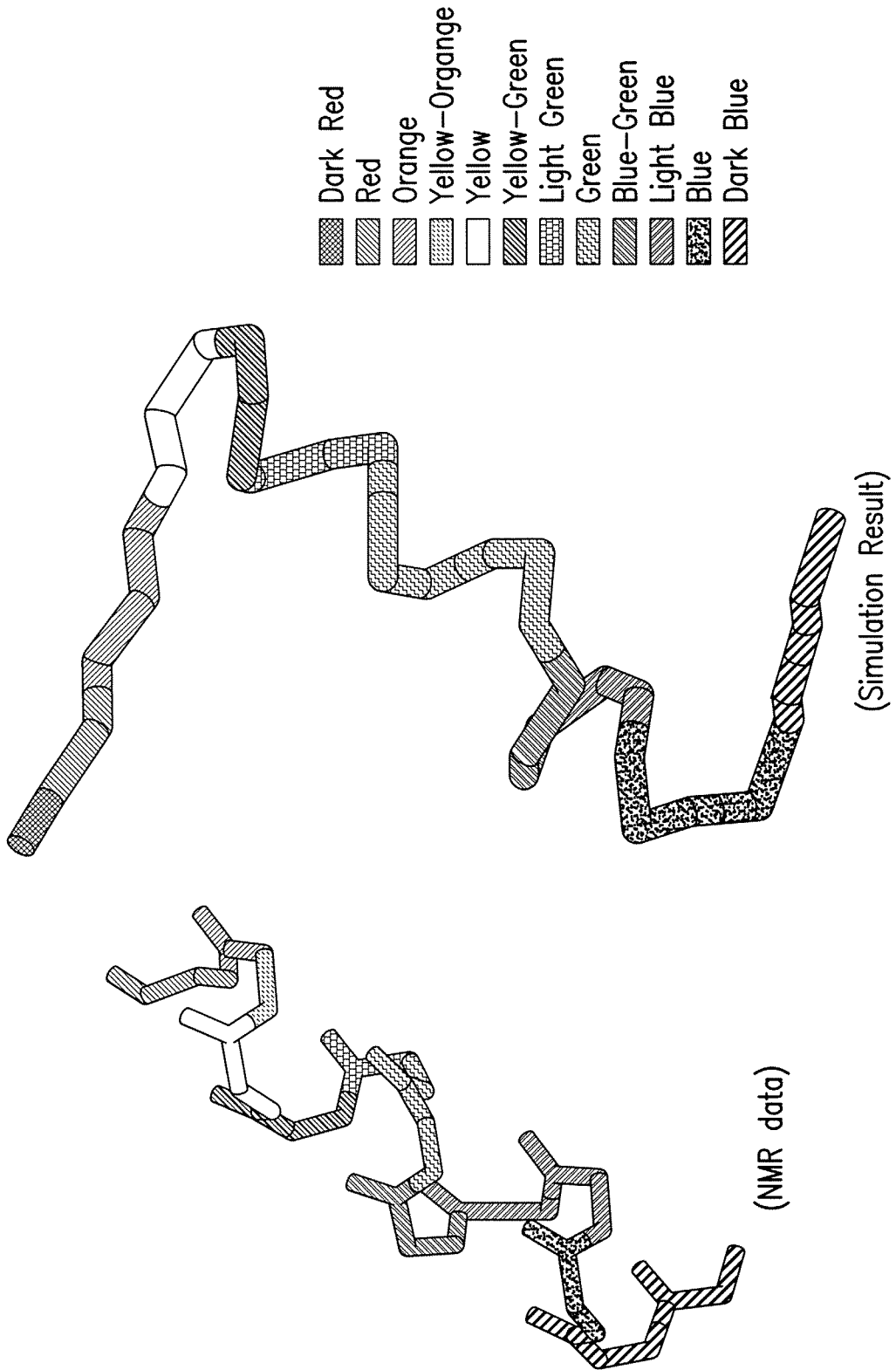


FIG.11

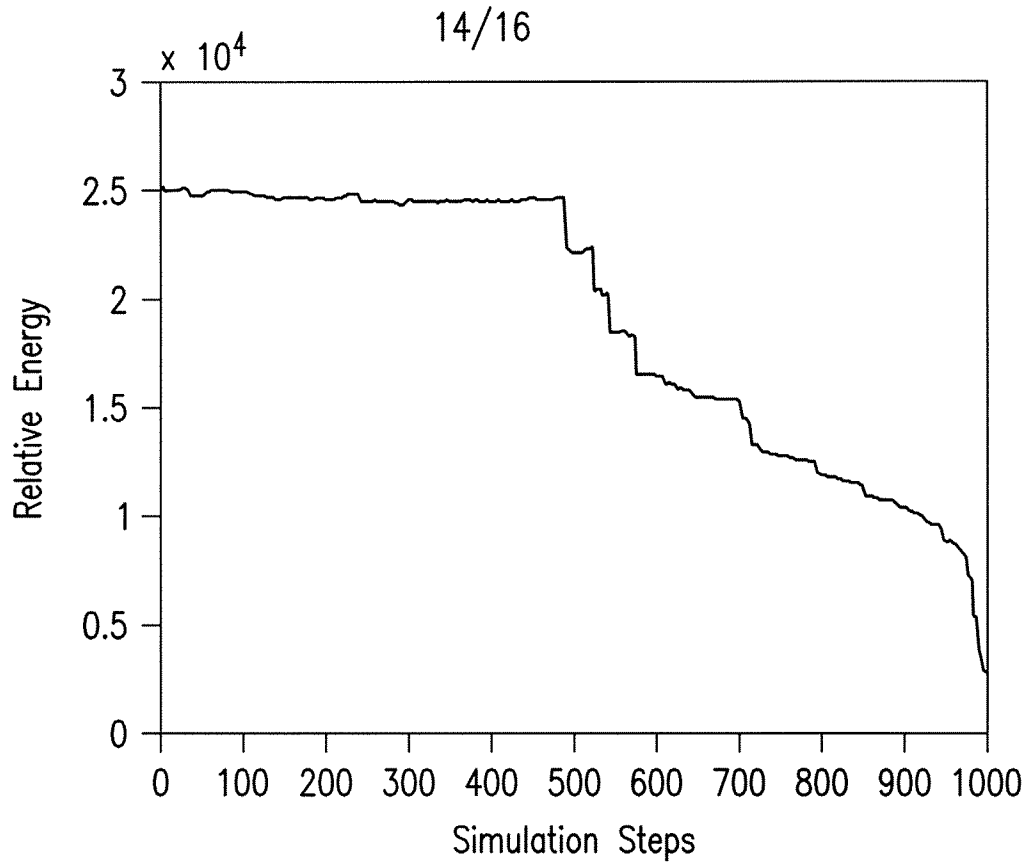


FIG.12A

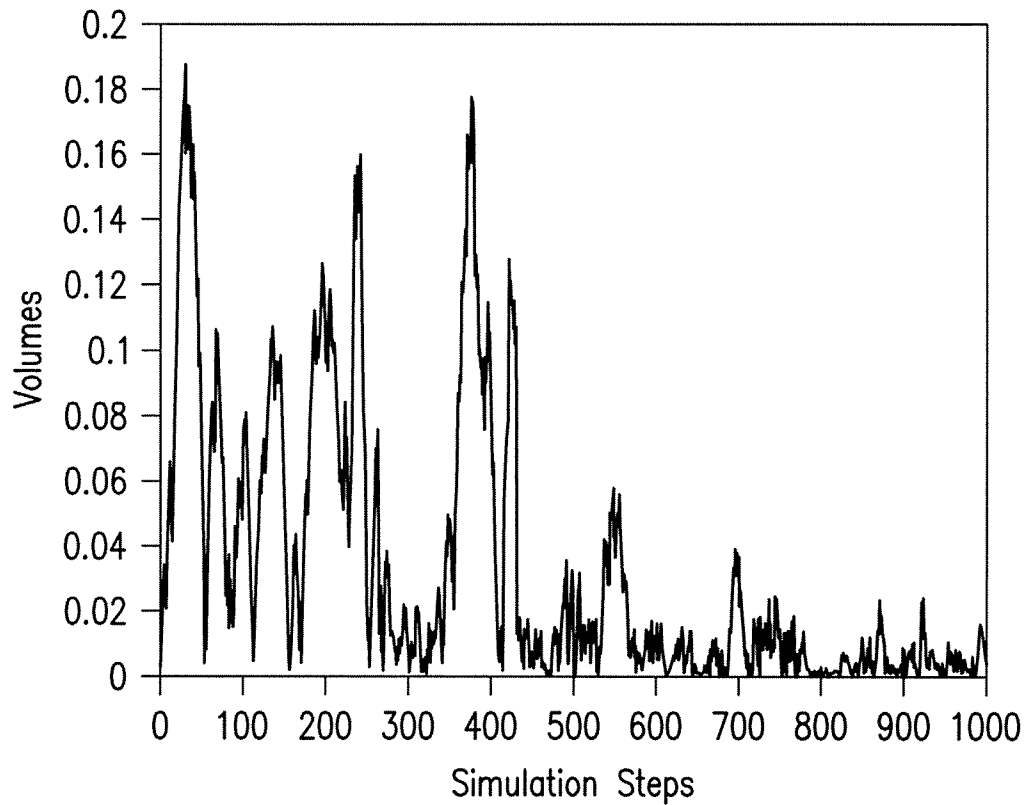


FIG.12B

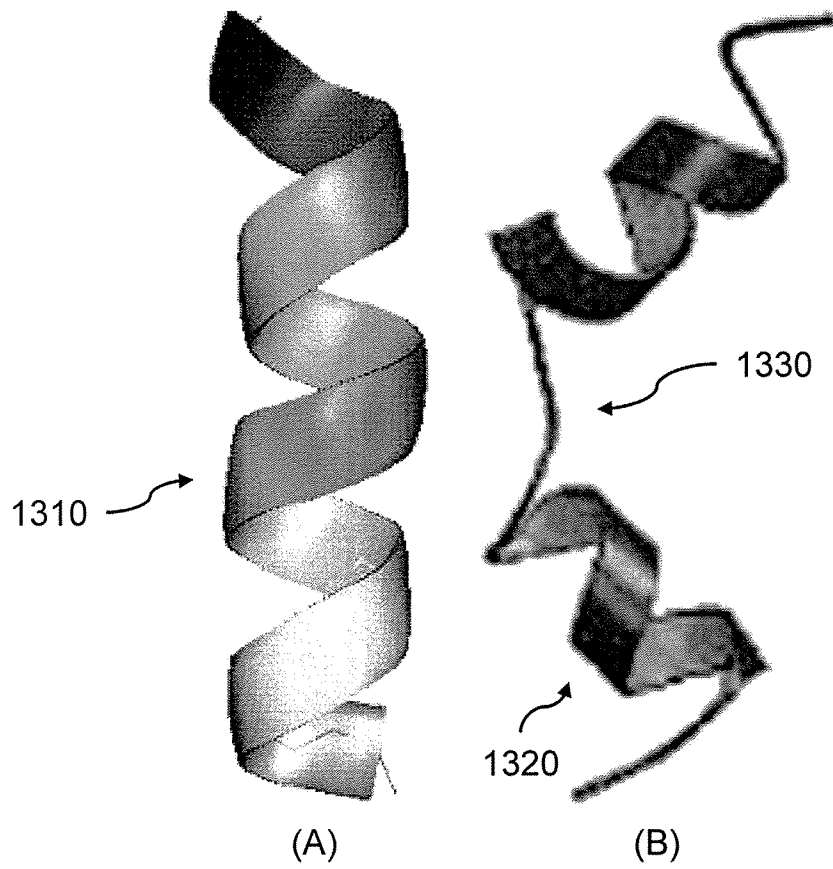


FIG.13

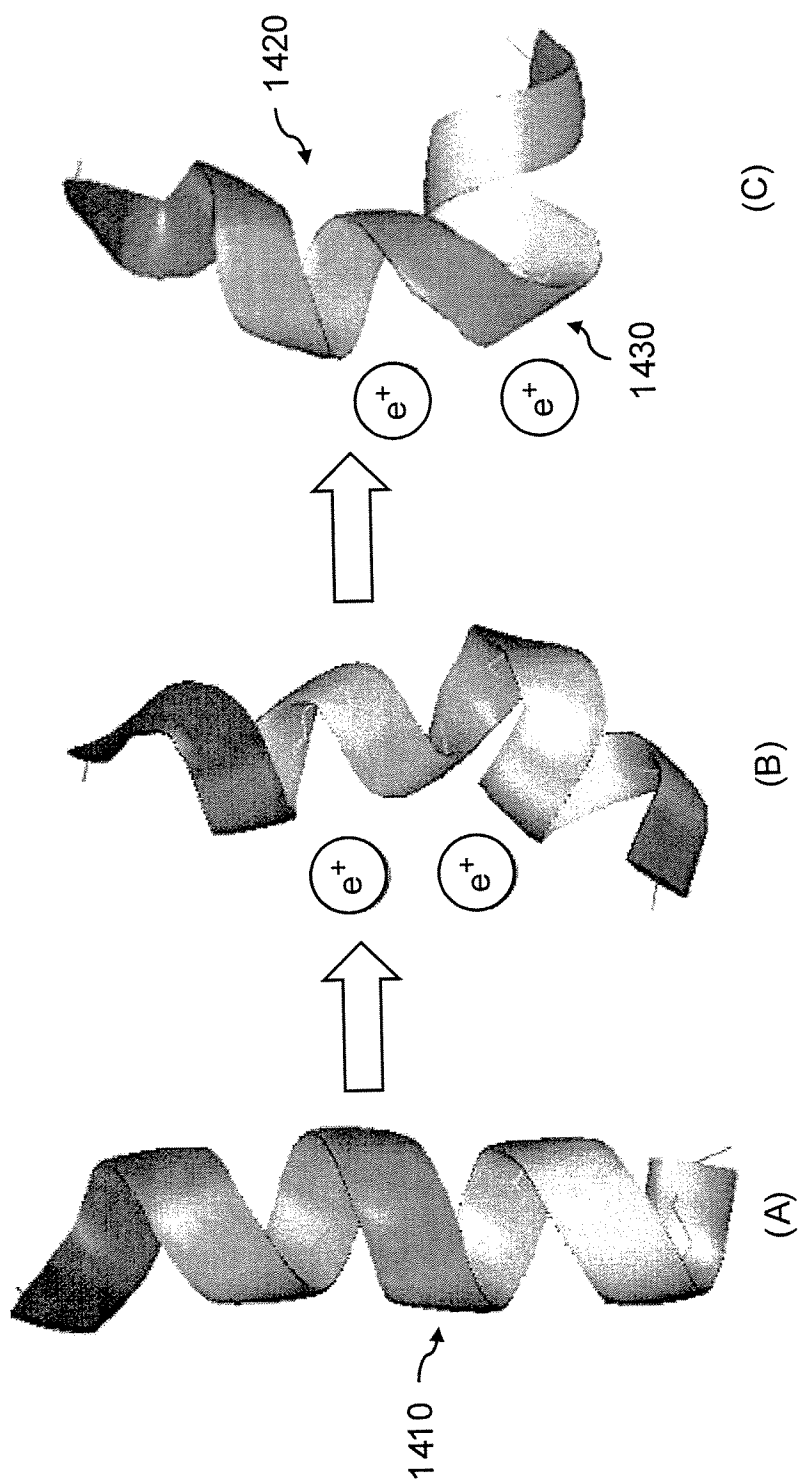


FIG.14