

# A Model for Protein Folding and Bio-catalysis

M. Pitkänen<sup>1</sup>, April 20, 2008

<sup>1</sup> Department of Physical Sciences, High Energy Physics Division,  
PL 64, FIN-00014, University of Helsinki, Finland.  
matpitka@rock.helsinki.fi, <http://www.physics.helsinki.fi/~matpitka/>.  
Recent address: Puutarhurinkatu 10,10960, Hanko, Finland.

## Contents

|          |                                                                                     |           |
|----------|-------------------------------------------------------------------------------------|-----------|
| <b>1</b> | <b>Introduction</b>                                                                 | <b>2</b>  |
| <b>2</b> | <b>Folding code and its generalization</b>                                          | <b>3</b>  |
| 2.1      | First trial . . . . .                                                               | 4         |
| 2.2      | Second trial . . . . .                                                              | 7         |
| 2.3      | How DNA nucleotides are connected with the hydrophilic<br>ends of lipids? . . . . . | 11        |
| 2.4      | Third trial . . . . .                                                               | 12        |
| 2.5      | 4-D spin glass energy landscape and code of catalytic action .                      | 20        |
| <b>3</b> | <b>A simple quantitative model for protein folding and catalyst<br/>action</b>      | <b>21</b> |
| 3.1      | The model . . . . .                                                                 | 22        |
| 3.2      | Some general consequences . . . . .                                                 | 24        |
| 3.3      | Model for helical structures . . . . .                                              | 25        |
| 3.4      | Model for $\beta$ sheets . . . . .                                                  | 28        |
| 3.5      | Model for protein-protein binding sites . . . . .                                   | 29        |
| <b>4</b> | <b>Some simple tests for the flux tube model of protein folding</b>                 | <b>30</b> |
| 4.1      | The test . . . . .                                                                  | 31        |
| 4.2      | About the interpretation of the results . . . . .                                   | 32        |
| <b>5</b> | <b>Appendix: Data about the monomers A and B of Fis pro-<br/>tein</b>               | <b>34</b> |

## Abstract

The model for the evolution of genetic code leads to the idea that the folding of proteins obeys a code inherited from the genetic code in the sense that aminoacid behaves like the conjugate  $Y_c$  of the middle nucleotide of the codon  $XYZ$  coding for it and that flux tubes connecting aminoacids to each other connect conjugate aminoacids behaving like  $Y$  and  $Y_c$ . Also catalyst action would reduce to effective base pairing in this picture. After some trials one ends up with a general conceptualization of the situation with the identification of wormhole magnetic flux tubes as correlates for attention at molecular level so that a direct connection with TGD inspired theory of consciousness emerges at quantitative level. This allows a far reaching generalization of the DNA as topological quantum computer paradigm and makes it much more detailed. The final outcome is very simple quantitative model for folding and catalyst action based on minimization of energy and consistent with basic experimental facts as well as general ideas.

## 1 Introduction

The model for the evolution of genetic code leads [L6] to the idea that the folding of proteins obeys a code inherited from genetic code in the sense that aminoacid behaves like the conjugate  $Y_c$  of the middle nucleotide of the codon  $XYZ$  coding for it and that flux tubes connecting aminoacids to each other connect conjugate aminoacids behaving like  $Y$  and  $Y_c$ . This conjugation corresponds to hydrophilic-hydrophobic dichotomy. Also catalyst action would reduce to effective base pairing in this picture chemically and at the level of quarks associated with the flux tube to matter antimatter conjugation.

After some trials one ends up with a general conceptualization of the situation with the identification of wormhole magnetic flux tubes as correlates for attention at molecular level so that a direct connection with TGD inspired theory of consciousness emerges at quantitative level. This allows a far reaching generalization of the DNA as topological quantum computer paradigm and makes it much more detailed. The final outcome is very simple quantitative model for both protein folding and catalyst action based on minimization of energy and consistent with basic experimental facts as well as general ideas.

The model is based on the assumption that the contribution of a flux tube connecting aminoacid and its conjugate to the potential energy depends only the distance between the molecules in question. The extremals of the total interaction energy are same for any choice of the potential and

only the absolute minimum of the interaction energy depends on the choice of the potential. The simplest potential corresponds to the negative of the harmonic oscillator potential in accordance with the idea of quantum criticality and the fact that hydrophilic and hydrophobic residues tend to have large distance. Also the contribution of the interactions between neighboring aminoacids are expected to be present but are neglected in the simplest model. The comparison of the predictions of the simplest model with the folding data about some proteins turns out to be encouraging but further work is needed to see whether the model really works.

Several persons have helped me in writing this chapter. I want to express my gratitude to Ulla Mattfolk for informing about the idea of protein folding code and to Dale Trenary for interesting discussions, for suggesting proteins which could allow to test the model and for providing concrete help in loading data help from protein data bank. Also I want to thank Timo Immonen for discussions and for loaning the excellent book "Proteins: Structures and Molecular Properties" of Creighton. I am also grateful for Pekka Rapinoja for writing the program transforming protein data file to a form readable by MATLAB.

## 2 Folding code and its generalization

Protein folding code is something which is expected to exist but is not understood [1]. This inspired a work which led to three trials for the folding code. Also a natural generalization to a code for catalysis emerged.

a) The first trial for the folding code relies on the assumption that free aminoacid behaves like the conjugate of the dinucleotide of the codon  $XY$  coding for it. This would hold true for both free aminoacids and those contained by proteins.

b) Learning of some basic facts about proteins and their interactions [4] inspired the hypothesis that the aminoacid in the interior of aminoacid sequence behaves like the conjugate of the nucleotide  $Y$  of the codon  $XYZ$  coding for it. For the second trial free aminoacid was assumed to behave like the conjugate of  $XYZ$ .

c) Third trial differs from the second trial in that free aminoacid would behave like the conjugate of  $XY$ . This option seems to be most realistic one.

## 2.1 First trial

The prebiotic 2-code assigning to RNA dinucleotides aminoacids as catalysts for the reaction attaching RNA dinucleotide to RNA polymer (this is discussed in detail in[L6]) might define also the folding code. This code would specify also what kind of catalytic reactions can occur between proteins and would dictate the catalytic stereo chemistry - or rather its changes - to a high degree.

a) The cautious working hypothesis of the first trial is that living aminoacids are braided in the sense that from a given aminoacid (free or contained in protein) there emanates two braid strands represented by "wormhole" magnetic flux tubes - let as call them simply threads. The threads are colored in the sense that they carry four different colors so that one has 16 color pairs: the color is specified in terms of quarks and antiquarks in a manner that I have explained earlier. The colors are in one-one correspondence with DNA nucleotides  $A, T, G, C$ : this hypothesis has led to a quite variety of predictions already shown to be correct [L5]. The colors of aminoacid threads are determined completely by the dinucleotide  $XY$  of DNA codon  $XYZ$  coding for the aminoacid.  $X$  codes for the precursor of given aminoacid and would begin from the part of aminoacid common to all four aminoacids associated with the same dicodon  $XY$ . The thread corresponding to  $Y$  would begin from the residue of these 4 aminoacids.

b) If the aminoacid is coded by more than 4 codons, it can have two different colored thread pairs. Ser, arg, and leu are this kind of aminoacids in case of the nuclear genetic code. One can say that aminoacid "remembers" which was the pair  $XY$  for the DNA codon coding it. The two variants of ser will be denoted by  $ser_1$  and  $ser_2$ , same for leu and arg.

c) The thread pair can connect aminoacid to another aminoacid, DNA triplet or RNA dinucleotide or perhaps even more general braided biomolecule (say precursor of aminoacid). Given aminoacid is effectively equivalent to the dicodon  $XY$  appearing in the codons coding for it and the basic step of the biocatalytic reactions would be analogous to base pairing. Genes would not code only for the aminoacids but also for their stereochemistry. In a well-defined sense aminoacids and DNA and RNA dinucleotides would form a social network in which two members are friends if they correspond to dicodon  $XY$  and its conjugate  $X_cY_c$ . One might also speak about molecular sex. The potential companions of the aminoacid associated with dicodon  $XY$  would be aminoacids associated with dicodon  $X_cY_c$ . Also these DNA and RNA dicodons would be potential companions of the aminoacid. An open question is whether aminoacid can attach to any dicodon in DNA and

RNA sequence or only to the dicodon part  $XY$  of codon  $XYZ$ : if so also DNA rather than only mRNA and tRNA could contain information about 3-codon decomposition of gene.

d) The phase transitions reducing Planck constant for the magnetic flux tubes defining the threads could bring aminoacid and its conjugate to the vicinity of each other. If the folding involves phase transitions reducing Planck constant, this makes possible to make a list about possible self contacts of protein once one knows the amino-acid sequence. In the case of catalytic reactions involving aminoacids, RNA, and DNA similar list about possible contact points between reactants can be given. That biocatalysis would reduce to symbolic dynamics based on gluing together of pieces of text and cotext would have extremely far reaching implications. For instance, aminoacid sequences attaching to DNA and catalyzing various kinds of processes should obey these rules.

e) In terms of the code table the rule would be that the companions of a given aminoacid are found by going in the code table two units up or down and two units right or left so that one remains inside the code table (see the table below). A couple of examples about this bio-molecular social network are in order.

i) The conjugates (companions) of Phe and  $Leu_1$  are Asn and Lys and the conjugates of  $Leu_2$  are Asp and Glu.  $Arg_1$  and Ala conjugates as also Gly and Glu. The conjugates of Thr are Sys and Trp.

ii) The conjugates of  $Ser_1$  are  $Ser_2$  and  $Arg_2$ . Ser is its own conjugate and thus a completely exceptional aminoacid.

iii) Ile, met (which serves as starting aminoacid) and thr have the formal aminoacid associated with stop codons as a conjugate. Whether this has some physical meaning remains open.

Consider now objections against the proposal.

a) If the magnetic flux tubes connecting nucleotide and conjugate correlates strongly with base pairing, then also aminoacid sequence and its various conjugates could form analogs of DNA double strands such that the residues of the paired aminoacids are hydrogen bonded. In the case of  $\alpha$  helix and  $\beta$  sheet [2] this kind of mechanism is not involved since the hydrogen bonds are associated with the non-varying  $H_2N - (CH) - C = O$  part of the paired aminoacids and there are no selection rules telling which residues can be paired. One must carefully distinguish between ordinary chemistry and the dynamical selection rules coming from the proposal. It would be the possible changes of tertiary and quaternary structures of proteins about which the hypothesis can possibly say something. The interactions could also stabilize the secondary structure (like alpha helices and beta sheets)

created by "ordinary" chemical interactions.

b) Gly is an aminoacid for which one has  $R = H$ . The naive expectation that the magnetic flux tube pair should end up to hydrogen atom looks somewhat strange. One cannot avoid the question whether also water could be living in the sense that the hydrogen atoms of water molecules can be connected to biomolecules so that the phase transitions changing Planck constant could be an essential part of hydrophobic and hydrophilic quantum dynamics.

c) The presence of water is a key determinant in protein folding so that if the code is important for the folding, it must relate to the possibility that hydrophobic and hydrophilic interactions induce changes of the Planck constant for magnetic flux tubes. From the table of the side chain properties of aminoacids [3] one finds that aminoacids with  $Y = A, G$  are hydrophilic and polar so that hydrophily would correspond to quark matter and hydrophobia for quark antimatter. Quark antimatter would tend to be at protein interior surface and quark matter at protein exterior surface. The same would hold true for protein and its conjugate and the formation of nearby contacts would not be frequent. Lengthening of the magnetic flux tubes and thus an increase rather than reduction of Planck constant would be favored. The notion of molecular sex would suggest that partners tend to be very near to each other. Of course, the correlation between partners might be enough and temporary reduction of Planck constant might correspond to molecular sex in more concrete sense.

|   | T                               | C                                                         | A                              | G                                   |   |
|---|---------------------------------|-----------------------------------------------------------|--------------------------------|-------------------------------------|---|
| T | phe → {asn, lys}                | ser <sub>1</sub> → {ser <sub>2</sub> , arg <sub>2</sub> } | tyr → {ile, met}               | cys → thr                           | T |
|   | phe → {asn, lys}                | ser <sub>1</sub> → {ser <sub>2</sub> , arg <sub>2</sub> } | tyr → {ile, met}               | cys → thr                           | C |
|   | leu <sub>1</sub> → {asn, lys}   | ser <sub>1</sub> → {ser <sub>2</sub> , arg <sub>2</sub> } | stop <sub>1</sub> → {ile, met} | stop <sub>2</sub> → thr             | A |
|   | leu <sub>1</sub> → {asn, lys}   | ser <sub>1</sub> → {ser <sub>2</sub> , arg <sub>2</sub> } | stop → {ile, met}              | trp → thr                           | G |
| C | leu <sub>2</sub> → {asp, glu}   | pro → gly                                                 | his → val                      | arg <sub>1</sub> → ala              | T |
|   | leu <sub>2</sub> → {asp, glu}   | pro → gly                                                 | his → val                      | arg <sub>1</sub> → ala              | C |
|   | leu <sub>2</sub> → {asp, glu}   | pro → gly                                                 | gln → val                      | arg <sub>1</sub> → ala              | A |
|   | leu <sub>2</sub> → {asp, glu}   | pro → gly                                                 | gln → val                      | arg <sub>1</sub> → ala              | G |
| A | ile → {tyr, stop <sub>1</sub> } | thr → {cys, stop <sub>2</sub> , trp}                      | asn → {phe, leu <sub>1</sub> } | ser <sub>2</sub> → ser <sub>1</sub> | T |
|   | ile → {tyr, stop <sub>1</sub> } | thr → {cys, stop <sub>2</sub> , trp}                      | asn → {phe, leu <sub>1</sub> } | ser <sub>2</sub> → ser <sub>1</sub> | C |
|   | ile → {tyr, stop <sub>1</sub> } | thr → {cys, stop <sub>2</sub> , trp}                      | lys → {phe, leu <sub>1</sub> } | arg <sub>2</sub> → ser <sub>1</sub> | A |
|   | met → {tyr, stop <sub>1</sub> } | thr → {cys, stop <sub>2</sub> , trp}                      | lys → {phe, leu <sub>1</sub> } | arg <sub>2</sub> → ser <sub>1</sub> | G |
| G | val → {his, gln}                | ala → arg <sub>1</sub>                                    | asp → leu <sub>2</sub>         | gly → pro                           | T |
|   | val → {his, gln}                | ala → arg <sub>1</sub>                                    | asp → leu <sub>2</sub>         | gly → pro                           | C |
|   | val → {his, gln}                | ala → arg <sub>1</sub>                                    | glu → leu <sub>2</sub>         | gly → pro                           | A |
|   | val → {his, gln}                | ala → arg <sub>1</sub>                                    | glu → leu <sub>2</sub>         | gly → pro                           | G |

Table 6. Genetic code and the first proposal for the folding code.

## 2.2 Second trial

There exists a wonderful book "Proteins: Structures and Molecular Properties" by Thomas E. Creighton published 1993 [4] and I am grateful for Timo Immonen for loaning this book to me. In the following I freely refer to the general facts discussed in this book rather than referring separately to every detail. From this book I learned that the first guess for the code of catalysis was wrong.

The reading inspired the hypothesis that *free* aminoacid should behave like the conjugate of the DNA codon  $XYZ$  rather than only  $XY$ . This conclusion seems to be wrong but led to the prediction that the aminoacid inside aminoacid sequence effectively reduces to  $Y$  since the formation of the peptide bonds by the elimination of water molecule and formation of  $\text{NH}-\text{O}=\text{O}$  hydrogen bonds effectively eliminates  $X$  and  $Z$ . This prediction seems to be realistic. The ends of aminoacid are predicted to behave like dicodons  $XY$  and  $YZ$  which could explain their special role in bio-catalysis. Only aminoacids for which  $Y$  corresponds to quarks (not antiquarks) can form hydrogen bonds so that hydrophilic-hydrophobic dichotomy corresponds to a strong matter antimatter asymmetry at quark level.

### 1. Matter antimatter asymmetry at the level of interactions of aminoacids

The first thing that I learned was that in the case of aminoacid belonging to protein interior second nucleotide  $Y$  in the codon  $XYZ$  coding for aminoacid is what matters. Only  $Y = A, G$  aminoacid residue can form hydrogen bonds and is hydrophilic and thus interacts strongly with water and DNA and RNA. In  $T, C$  case the formation of hydrogen bonds is impossible or rare (ser,thr). In their interactions with water these aminoacids are passive, or rather-avoid water- and tend to interact with each other. This division is fundamental for the understanding of the interactions of aminoacids. The division of aminoacids to hydrophobic *resp.* non-hydrophobic ones corresponds to the assignment of quarks to  $A$  and  $G$  and antiquarks to  $T$  and  $C$  so that strong matter antimatter asymmetry is in question. Similar asymmetry appears in cosmology: in TGD Universe antimatter resides inside cosmic strings in the interior of big voids containing matter as galaxies at their boundaries so that one can understand why antimatter is not visible.

### 2. Flux tubes can connect with all electronegative atoms

Also a plausible answer to the question which atoms can be connected by flux tubes emerges.

a) The model for dinucleotide precursor code [L6] involves precursors

for which 3 precursors contain only oxygen ions or double bonded oxygens. The only possible conclusion is that oxygen can connect to any DNA letter (quark or antiquark) and that first letter-precursor correlation is a selection of the most probable alternative. Also in water oxygen atoms should form flux tube contacts with each other and aminoacids and DNA. Also nitrogen atoms might form similar flux tube connections and this was assumed in the model. Same would apply to sulphur appearing in met and tyr and to electronegative atoms in general.

b) The guess that the presence of the flux tube would be a necessary prerequisite for the hydrogen bond formation is wrong. Indeed, hydrogen bonds are formed between polar groups of hydrophilic aminoacids. The di-sulphur associated with cys-cys pairs play a fundamental role in protein folding. This bond is not allowed by the generalized base pairing rule.

c) Hydrophobic aminoacids could connect with the oxygen in water by flux tubes but they could not form hydrogen bonds. The phase transition increasing  $\hbar$  would allow them to increase their distance from water molecules in a controlled manner. This could be essential for folding and make possible the formation of pockets connected by flux tubes of large  $\hbar$  to water. In quantum models for evolution of consciousness these pockets are believed to play a prominent role. Molecular sex would thus mean correlation tending to keep partners at large distance except when  $\hbar$  reducing phase transition occurs.

### *3. What can one learn from the formation of alpha helices and beta sheets?*

The formation of peptide bonds by the elimination of  $H_2O$  molecules and generation of hydrogen bonds between NH and O= is an essential step in the formation of alpha helices and beta sheets. Second observation is that aminoacids decompose naturally into three parts corresponding to O=COH, R, and  $NH_2$ . This suggests that aminoacid actually corresponds to the entire DNA codon  $XYZ$  coding for it. OH could correspond to  $Z$ , R to  $Y$ , and  $NH_2$  to  $X$ . In the case of pro the residue binds to  $N$  so that a closed cycle is formed and it is  $NH$  which would be correspond to  $Z$  in this case. In case of gly R is simply hydrogen atom and would be connected by flux tube to DNA. Both of these observations could be used as objections against this picture.

In the formation of peptide bond the flux tube connecting to COH and thus to  $Z$  would be taken by the water molecule created in the formation of peptide bond leaving only  $XY$ . The first flux tube would connect HN and O= so that  $X$  would pair with  $X_c$  assignable to O. There are no problems



with the formation of bond if O= can correspond to any code letter as in the case of water. Water would correspond to matter antimatter symmetric phase and an interesting question is what counterpart this phase could have in cosmology (bosonic matter?).

The aminoacid inside protein would effectively behave like  $Y_c$  in the effective base pairing. Depending on whether it corresponds to quark or antiquark, aminoacid would be hydrophilic or hydrophobic- or rather - able to form hydrogen bonds or not. Since hydrophobic aminoacids cannot form hydrogen bonds, the formation of these residue pairs would be inhibited. The hydrophilic and hydrophobic residues could tend to avoid each other and the phase transitions increasing Planck constant would make this possible. It must be emphasized that this brings in strong long range correlation between the dynamics of the aminoacid residues belonging to the first and third (second and fourth) column of the code table.

Hydrophilic aminoacids would form hydrogen bonds which each other and with DNA and RNA. In catalytic binding sites this kind of hydrogen bonds are formed between polar groups: also hydrogen bonds with water are formed and they tend to neutralize possible static charges. Ser ( $UCZ$ ) and thr ( $ACZ$ ) are the only effectively hydrophobic aminoacids containing OH group (and thus strictly speaking amphiphilic). Perhaps it is not an accident thr the codon ACC coding for thr appears in the stem of tRNA containing aminoacid. Ser and thr are indeed able to form hydrogen bonds with hydrophilic aminoacids and the prediction is that these aminoacids have form  $XGZ$  belonging to the last column of the code table. According to [4] there are however very few biochemical reactions of this kind useful for proteins. Ser is exceptional in that it is predicted to be able to form flux tubes connecting ser<sub>1</sub> coded by  $TCZ$  with ser<sub>2</sub> coded by  $AGZ$ ,  $Z = T, C$ . The OH group of ser can be seen as a correlate for this property.

The aminoacids at the ends of the polymer are predicted to behave effectively like dinucleotides. The aminoacid coded by XYZ would base pair like  $X_c Y_c$  if in the beginning of polymer and to  $Y_c Z_c$  if at the end of polymer. These nucleotides should have very special selective role in DNA-aminoacid and RNA-aminoacid interactions. Remarkably, it is known that the cutting of COOH and NH<sub>2</sub> away from the end of polymer in general makes protein folding impossible (also mutations can affect dramatically folding). The first nucleotide of protein is usually met containing sulphur and the conjugation associates met with stop and tyr codons. The association of met with stop is indeed natural for the free NH<sub>2</sub> of met having no hydrogen bond in the beginning of the sequence.

According to [4], the binding sites of catalyst and ligand in the reac-

tion complex are conjugates both geometrically and physically. It would be nice to have a concrete representation of this conjugacy in terms of the genetic code. Geometric conjugacy is easy to understand in terms of the lock and key picture and physical conjugacy means among other things that hydrophilic aminoacids that behave as acids *resp.* bases combine with other in the interface. This does not correspond to  $Y - Y_c$  pairing. A detailed quantitative model for binding leads to a model of what happens in interface. The interfaces can be thought of as cutting protein along its interior: in center there are hydrophobic aminoacids and in periphery hydrophilic ones. The flux tubes must connect periphery of A (B) to center of B (A).

The strong correlation between RNA dinucleotide and aminoacid in the case of tRNA conforms with this picture. The third flux tube associated with the aminoacid would connect with the third nucleotide after the transition to RNA-aminoacid era. During RNA era tRNA<sub>2</sub> would have connected the O=C-OH part of the aminoacid to water molecule.

#### 4. Interactions with DNA

Also in the interactions with DNA and RNA the aminoacid in the interior of the sequence would base-pair" like  $Y_c$ . The original idea about molecular sex would be modified. Either one accepts that aminoacid and its conjugate tend to be far away but correlated and come near to each other only in phase transitions reducing Planck constant, or that the companion of the hydrophilic aminoacid is identified as DNA nucleotide in general. Hydrophobic aminoacids would behave like hermits. The generic contacts with DNA would be contacts with single nucleotide and there would be 4 different basic contacts. Aminoacids are indeed known to form contacts with single nucleotide. Hydrophilic contacts would be favored and hydrophobic contacts avoided so that again  $Y = A, G$  aminoacids would play at the outer boundary of DNA would play the active role. The aminoacids inside a given column of the code table would interact in very much the same manner with DNA nucleotides as far as formation of hydrogen bonds is considered. The terminals of the protein polymer are predicted to behave like  $X_c Y_c$  *resp.*  $Y_c Z_c$  if the corresponding codon is  $XYZ$ . Again only hydrophilic codons are expected to be able to form hydrogen bonds. N-terminal is usually met and met and should avoid DNA.

#### 5. Interactions of proteins with ions and electrons

Proteins interact also with electrons and ions. Typical process are the addition or removal of proton, electron, ion such  $Ca^{++}$ , or molecule such as  $O_2$ . These interactions are not well understood. For instance, the inter-

actions involve the transfer of electrons between ligand protein and protein inducing oxidation (electron is given), reduction (electron is received) or redox reaction (both reduction and oxidation take place). In metabolism redox process is central. These reactions are reversible and it is difficult to understand how electrons are able to make their long journey from the interior of the ligand so fast and avoiding dissipative effects. The formation of cyclotron Bose-Einstein condensates of bosonic ions and electronic Cooper pair condensates at the magnetic flux tubes connecting ligand and protein might provide the solution of the mystery. Note that the new nuclear physics predicted by TGD predicts nuclei which can have anomalous em charge associated with the color fluxtubes connecting nucleons to nuclear string so that fermionic ions  $Na^+$ ,  $Cl^-$ ,  $K^+$  could have exotic bosonic counterparts.

### 2.3 How DNA nucleotides are connected with the hydrophilic ends of lipids?

The starting point of all these developments was the model for DNA as a topological quantum computer (tqc) described in the earlier postings. It was assumed that braid strands defined by "wormhole magnetic" flux tubes join nucleotides to lipids and can continue through the nuclear or cell membrane but are split during tqc. The hydrophilic ends of lipids attach to water molecules and self-organization patterns for the water flow in gel phase induce a 2-D flow in the lipid layer which is liquid crystal defining tqc programs at the classical level as braidings. The flow indeed induces braiding if one assumes that during topological computation the connection through the cell membrane is split and reconnected after the halting of tqc.

The challenge is to understand microscopically how the flux tube joins DNA nucleotide to the phospholipid [5]. Certainly the points at which the flux tubes attach should be completely standard plugs and the formation of polypeptide bonds is an excellent guide line here. Recall that phospholipid, the tqc dancer, has two hydrophobic legs and head. Each leg has at the hydrophilic end  $O=C-O-C$  part joining it to glyceride connected to monophosphate group in turn connected to a hydrophilic residue R. The most often appearing residues are serine, inositol, ethanolamine, and choline. Only three of these appear in large quantities and there is asymmetry between cell exterior and interior.

Let us denote by  $=O_1$  and  $=O_2$  the two oxygens (maybe analogs of right and left hemispheres!) in question. The proposal is that DNA nucleotide and  $=O_1$  are connected by a flux tube: the asymmetry between right and left lipid legs should determine which of the legs is "left leg" and which  $O=$

is the "left brain hemisphere".  $=O_2$ , the "holistic right brain hemisphere", connects in turn to the flux tube coming from the other symmetrically situated  $=O_2$  at the outer surface of the second lipid layer. Besides this  $=O_1$  and  $=O_2$  are connected by a flux tube serving as switch on both sides of the membrane.

The assumption that two flux tubes enter to  $=O$  is consistent with the fact that two hydrogen bonds to  $=O$  are possible. During tqc this short flux tube is split or disappears. The lipid residue R couples with the flow of the liquid in gel phase. Since  $=O$  is in question the quark or antiquark at the end can correspond to the DNA nucleotide in question. The necessary complete correlation between quark and antiquark charges at the ends of flux tubes associated with  $=O_1$  and  $=O_2$  can be understood as being due to the minimization of Coulomb interaction energy.

The phosphate groups associated with nucleotides of DNA strand contain also  $=O$ , which could act as a plug to which the flux tube from the nucleotide is attached.  $=O$  appears in biomolecules involved with varying functions such as signalling, control, and metabolism.  $=O$  might act as a universal plug to which flux tubes from electronegative atoms of information molecules can attach their flux tubes. This would also provide a concrete realization of the idea that information molecules (neurotransmitters, hormones) are analogous to links in Internet [M2]: they would not represent the information but establish a communication channel. The magnetic flux tube associated with the information molecule would connect it to another cell and by the join to  $=O$  plug having flux tube to another cell, say to its nucleus, would create a communication or control channel.

## 2.4 Third trial

Biochemistry represents extremely complex and refined choreography. It is hard to believe that this reduces to a mere unconscious and actually apparent fight for chemical survival. In TGD Universe consciousness would be involved even at the molecular level and magnetic body would be the choreographer whose dance would induce the molecular activities. This picture combined with the idea of standard plugs and terminals at which flux tubes end, leads to a third trial to understand catalytic code.

The third trial differs from the second trial in that the letters  $X, Y, Z$  of the codon  $XYZ$  coding for the aminoacid do not correspond to  $COOH$ , residy R, and  $NH_2$  ( $NH$  in case of pro) group so that free aminoacid would behave like conjugate of  $XYZ$ . Rather, free aminoacid behaves like  $XY$  as in the first trial and  $X$  and  $Y$  correspond to flux tubes ending at  $OH$  and

=O in COOH group. For the new option all - not only alpha helical and beta sheeted - aminoacids in the interior of the aminoacid sequence behave like the conjugate of letter  $Y$  for the codon  $XYZ$  coding for the aminoacid. The new model predicts that DNA, mRNA, tRNA, and aminoacids are in general connected by braid strands and provides a detailed picture about the role of braidings in transcription and translation. The topological dynamics of the magnetic body, its motor activities, would induce catalytic dynamics.

*1. Flux tubes as a correlate for directed attention*

Molecular survival is the standard candidate for the fundamental variational principle motivating the molecular intentional actions. There is entire hierarchy of selves and the survival at the higher level of hierarchy would force co-operation and altruistic behavior at the lower levels. One might hope that this hypothesis reduces to Negentropy Maximization Principle [H2], which states that the information contents of conscious experience is maximized. If this picture is accepted, the evolution of molecular system becomes analogous to the evolution of a society.

Directed attention is the basic aspect of consciousness and the natural guess would be that directed attention corresponds to the formation of magnetic flux tubes between subject and target. The directedness property requires some manner to order the subject and target.

a) The ordering by the values of Planck constant is what first comes in mind. The larger space-time sheet characterized by a larger value of Planck constant and thus at a higher level of evolutionary hierarchy would direct its attention to the smaller one.

b) Also the ordering by the value of p-adic prime characterizing the size scale of the space-time sheet could be considered but in this case directedness could be questioned.

c) Attention can be directed also to thoughts. Could this mean that attention is directed from real space-time sheets to p-adic space-time sheets for various values of primes but not vice versa? Or could the the direction be just the opposite at least in the intentional action transforming p-adic space-time sheet to real space-time sheet? Perhaps directions are opposite for cognition and intention.

The generation of wormhole magnetic flux tubes could be the correlate for the directed attention, not only at molecular level, but quite generally. Metaphorically, the strands of braid would be the light rays from the eyes of the perceiver to the target and their braiding would code the motions of the target to a topological quantum computation like activity and form a memory representation at least. The additional aspect of directed attention

would be the coloring of the braid strands, kind of coloring for the virtual light rays emerging from the eyes of the molecular observer. In the case of DNA this can induce a coloring of braid strands emerging from aminoacids and other molecules so that it would indeed become possible to assign to aminoacid the conjugate of the middle nucleotide of the codon  $XYZ$  coding for it.

Attention can be also redirected. For this process there is a very nice topological description as a reconnection of flux tubes. What happens is that flux tubes  $A \rightarrow B$  and  $C \rightarrow D$  fuse for a moment and become flux tubes  $A \rightarrow D$  and  $C \rightarrow B$ . This process is possible only if the strands have same color so that the values of the quark charges associated with  $A$  and  $B$  are same.

This kind of process can modify tqc programs. For instance, in the case of the flux tubes coming from nucleotides  $X$  and  $X_c$  and ending to the lipid layer this process means that  $X$  and  $X_c$  and corresponding lipids become connected and genome builds memory representation about this process via similar link. If proteins are connected with mRNA connected to DNA in this manner, this process would allow the formation of flux tubes between aminoacids of two proteins in such a manner that protein would inherit from DNA codon the color of the middle nucleotide and its interactions effectively reduce to base pairing.

DNA would have memory representation about molecular processes via these changing braiding topologies, and one could say that these molecular processes reflect the bodily motions of the magnetic body. Entire molecular dynamics of the organism could represent an enormous tqc induced by the motor activities of the magnetic body. At the level of sensory experience similar idea has been discussed earlier [H11]: out of body experiences (OBEs) and illusions such as train illusion could be understood in terms of motor action of magnetic body inducing virtual sensory percepts.

Attention can be also switched on and off. Here the structure of the lipid ends containing two nearby situated =O:s suggest the mechanism: the short flux tube connecting =O:s disappears. The minimization of Coulomb interaction energy at each end implies that re-appearance of the flux tubes creates a short flux tube with the original strand color.

## *2. Where do flux tubes begin from?*

The view about magnetic body as a controller of biological body using genome as a control tool suggests that DNA is to a high degree responsible for directed attention and other molecules as targets so that flux tubes emanate from DNA nucleotides. The reason would be that the aromatic

cycles of DNA correspond to larger value of Planck constant.

Some chemical or geometric property of DNA nucleotides or of DNA nucleotides of DNA strand could raise them to the role of subject. Aromatic cycle property correlates with the symmetries associated with large value of Planck constant and is the best candidate for this property. If this is accepted then also some aminoacid residues might act as subjects. Phe, His, Trp, Tyr contain aromatic cycle. The derivatives of Trp and Tyr act as neurotransmitters and His is extremely effective nucleophilic catalyst. This would make possible more specific catalytic mechanisms through the pairing of Phe, His, Trp, and Tyr with residues having flux tube terminals.

This raises the question about the physical interaction determining the color of the strand emerging from the aromatic cycle. The interaction energy of quark at the end of flux tube with the classical electromagnetic fields of nuclei and electrons of the ring should determine this. The wormhole contact containing quark/antiquark at the throat at space-time sheet containing nuclei and electrons could also delocalize inside the ring. One of the earliest hypothesis of TGD inspired model for living matter was that wormhole Bose-Einstein condensates could be crucial for understanding of the behavior of biomolecules [J5]. Wormhole throats with quark and anti-quark at their throats appear also in the model of high  $T_c$  superconductivity [J1]. As far as couplings are considered, these wormhole contacts are in many respects analogous to the so called axions predicted by some theories of elementary particle physics. The wormhole contact like property is by no means exceptional: all gauge bosons correspond to wormhole contacts in TGD Universe.

The only manner for the electronic space-time sheet to feed its electromagnetic gauge flux to larger space-time sheets using exactly two wormhole contacts is to use wormhole contacts with  $\bar{u}$  and  $d$  at their "upper" throat ( $T, G$ ). For proton one would have  $\bar{d}$  and  $u$  at their "upper" throat ( $A, C$ ). The presence of electron or proton at nucleotide space-time sheet near the end of flux tube might allow to understand the correlation. The transfer of electrons and protons between space-time sheets with different p-adic length scale is basic element of TGD based model of metabolism so that there might be some relation.

### *3. What aminoacids can act as plugs and terminals of flux tubes?*

Standardization constraint suggests that flux tubes are attached to standard plugs and terminals. The explicit study of various biological molecules and the role of water in biology suggests that =O serves as a plug to which flux arrives and from which it continues. The intuitive reason for the pro-

posal is that =O allows two hydrogen bonds. OH would in turn correspond to a terminal at which flux tube ends. One might be very naive and say that conscious biomolecules have learned the fundamental role of oxygen and water in the metabolism and become very attentive to the presence of =O and OH. =O appears in the residues of Asp, Glu, Asn, Gln. OH groups appear inside the residues of Asp, Glu and Ser, Thr.

It might not be very wise to restrict the molecular attention to only =O and OH and it is probably better to speak about probabilities for the flux tubes to attach to various kinds of terminals. Both SH and NH<sub>2</sub> are chemically like OH both these them could act as terminals of flux tubes: NH<sub>2</sub> (Asp, Gly, Glu, Arg) contain NH<sub>2</sub> and Cys contains SH.

#### *4. Directed attention generates memory representations and tqc like processes*

Directed attention induces braiding if the target is moving and changing its shape. This gives rise to a memory representation of the behavior of the object of attention and also to a tqc like process. A considerable generalization of tqc paradigm suggests itself.

Tqc could be induced by the braiding between DNA and lipids, DNA and proteins via folding processes, DNA RNA braiding and braiding between DNA and its conjugate, DNA and protein braiding. The outcome of tqc would be represented as the temporal patterns of biochemical concentrations and rates and there would be hierarchy of p-adic time scales and those associated with the dark matter hierarchy.

For instance, the protein content of lipid membranes is about 50 per cent and varies between 25-75 per cent so that protein folding and lipid flow could define tqc programs as self-organization patterns. The folding of protein is dynamical process: alpha helices are created and disappear in time scale of  $10^{-7}$  seconds and the side chains of protein can rotate.

The details of the tqc like process depend on what one assumes. The minimal scenario is deduced from the transcription and translation processes and from the condition that magnetic body keeps control or at least keeps book about what happens using genome as a tool. The picture would be essentially what one might obtain by applying a rough model for web in terms of nodes and links. The reader is encouraged to use paper and pencil to make the following description more illustrative.

a) mRNA and DNA must remain connected by flux tubes after transcription. The  $Y_c$  of mRNA codon could be connected to =O plug in the aminoacid of tRNA molecule and this to  $Y$  in tRNA anticodon so that one would have DNA-aminoacid-tRNA link.  $Z_c$  in mRNA would be connected



to  $Z$  in tRNA anticodon giving mRNA-tRNA link. OH in aminoacid would be connected to  $X$  in tRNA dicodon  $XY$  giving aminoacid-tRNA link.

b) When tRNA donates its aminoacid to the growing chain, the formation of the peptide bond separates one  $H_2O$  and the  $X$  connection to  $OH$  becomes a connection to water molecule so that one obtains tRNA- $H_2O$  link. DNA-mRNA-aminoacid-tRNA link with color  $Y$  is preserved. In depolymerization of mRNA  $H_2O$  molecule per aminoacid is used and the reverse change for linkings takes place.

c) The recombination process for two conjugate DNA-mRNA-aminoacid-tRNA links can transform the flux tubes in such manner that one obtains link between the =O:s of aminoacids  $A_1$  and  $A_2$  characterized by  $Y$  and  $Y_c$ . As proposed, this mechanism could be central in the enzyme substrate interaction. The process would pair tRNAs corresponding to  $Y$  and  $Y_c$  together to give DNA-mRNA-tRNA-tRNA-mRNA-DNA link providing a memory representation about aminoacid pairing  $A_1 - A_2$ . One can say that magnetic body creates with the mediation of the genome dynamical tqc programs to which much of the biomolecular activity reduces. Not all however, since two aminoacid pairs  $A_1 - A_2$  and  $A_3 - A_4$  can recombine to  $A_1 - A_4$  and  $A_3 - A_2$  without DNA knowing anything about it. Magnetic body would however know.

d) The constant part of the aminoacid inside aminoacid would behave like  $Y_c$  if aminoacid is coded by  $XYZ$  whereas the ends and the protein would behave like conjugates of dicodons  $XY$  (for second trial it would have behaved like conjugate of  $YZ$ ). If one assigns to the hydroxyl and amino groups of the residue the roles of object and subject also flux tubes connecting the residue groups become possible and protein does not behave like single nucleotide anymore although one can still say that everything reduces in a well-define sense to the genetic code.

##### *5. Introns and DNA-protein attachment*

An example is the situation in which protein acts as an enzyme attaching on DNA. Suppose that this process effectively reduces to a base pairing between aminoacid and DNA nucleotide. Protein can attach to any portion of DNA. The simplest interaction is the attachment to the gene coding for the aminoacid itself but much more general enzymatic interactions are possible. It must be however noticed that DNA sequence coding for given aminoacid sequences is considerably longer than aminoacid sequence: the sequence coding for 10 aminoacids is about 10 nm long whereas the corresponding straight aminoacid strand is about 4.7 nm long. It is known that DNA can change its conformation from strand during enzyme-DNA action

[4], and the contraction of DNA strand might make possible to have enzyme-DNA interaction involving fusion along several subsequent aminoacids. This kind of mechanism might work also in the case that attachment region corresponds to several exons. There is however no need to assume that subsequent aminoacids are form a contact with DNA.

One can of course ask whether genes containing introns tend to code for proteins which are used for topological quantum computations. Introns, perhaps the repeating sequences with no obvious function, would have at least this useful function but very probably much more useful ones too (they are now known to be transcribed to RNA and TGD suggest that language corresponds to intronic gene expression). The emergence of introns might be somewhat like the emergence of information society.

The foldings of proteins tend to be conserved in the evolution whereas primary structure can change quite a lot apart from some aminoacids critical for enzymatic action. This confirms with the effective base pairing interaction between aminoacids and DNA and would mean that DNA-aminoacid tqc programs are rather robust against mutations.

#### *6. Evolution and braidings*

The evolution at the molecular level corresponds to the emergence of increasingly complex molecules using as basic building blocks aminoacid chains and non-translated residues attached to them in the post-translational processing of the aminoacid chains. Also increasingly complex reaction paths emerge. Molecular survival and the competition for the metabolic resources at molecular level could be seen as the basic driving force of this evolution.

Typically, in the original situation the enzymes would have received the substrate molecules from the environment but sooner or later this would have become difficult. The solution would have been a synthesis of the substrate from simpler ingredients by starting from some precursor.

If molecules (with magnetic bodies included) are conscious entities able to direct attention, one can imagine that magnetic body controlling them with the mediation of genome and able to actively modify it, could help through modifications of the genome to create to the catalyst a binding site able to bind the precursor. Immune system is doing this very intensively. If the enzyme binding the precursor already exists, a combination of genes coding for the enzyme and the enzyme having the metabolites as ligands could allow to achieve this. All this would reduce to the motor activities of magnetic body, in particular reconnection of flux tubes, a kind of dance of Shiva. Genome would not be anymore a sequence of DNA developing through random mutations under selection pressures.

In this framework aminoacids would have appeared before their precursors and possessed some function in RNA world, say the catalysis of join of RNA<sub>2</sub> dinucleotides to the increasing chain as proposed in [L6]. Competition might have led to a situation in which RNA<sub>2</sub> learned to catalyze selectively the generation of aminoacids from much simpler precursors (three of the proposed precursors contain only C,=O, and O<sup>-</sup>) giving rise to positive feedback implying an exponential amplification of RNA and aminoacid populations. The reduced genetic code would have been present at two levels. The reader can decide whether this is a shortcoming of the model or a fundamental biochemical duality.

Can one make any clear cut predictions about preferred mutations?

a) Mutations are not expected to be always random point mutations but could be a result of a purposeful action of the magnetic body. Chemical similarity is expected to be conserved in good mutations. This is known to be the case. Allowed point mutations should conserve  $Y$ . Also bi-local mutations of codons with middle nucleotides forming  $Y - Y_c$  pairs and conserving this property might occur and could be crucial for the coherence of the organisms. As found, the formation of flux tube between aminoacids  $A_1$  and  $A_2$  induces a flux tube between nucleotides  $Y$  and  $Y_c$  at the corresponding genes. This flux tube could force the possibly intentional mutations to occur as simultaneous point mutations of the two genes conserving the conjugacy property and leaving thus braiding invariant.

Some examples are in order. Ala/Ser, Ser/Thr, Ile/Val/Leu, Asp/Glu do not change  $Y$ . Lys/Arg ( $A/G$ ), Tyr/Phe ( $A/U$ ), Gly/Ala ( $G/C$ ),... are also prevalent and one might hope that they correspond to binary mutations in some important cases.

b) Folding is known to be more conserved than aminoacid sequence [4]. Since folding is a collective property of gene, local chemistry might not be enough and the proposed non-local conservation laws might be needed. Two-point mutations would also correlate the mutations of the binding sites of protein and ligand. The prediction would be conserved  $Y - Y_c$  pairs in genes coding for protein and ligand and these pairs might allow to deduce the paired points. The paired nucleotides need not belong to the same strand since genes are evenly distributed between strand and its conjugate and characterized by A,G surplus. A strong form of conjugacy stating that paired genes belong to the strand and its conjugate sounds beautiful in the ears of mathematician at least and would be the mirror image for the mutual avoidance of quark matter and antimatter at protein level.

c) If the flux tubes can connect also side chains, the situation becomes more complex. There is a temptation to think that these flux tubes would

connect only the nearby aminoacids of the same peptide and do not therefore affect the large scale dynamics of folding. This would be the case if the value of Planck constant associated with these flux tubes is smaller than for the flux tubes connecting aminoacids as basic units. If flux tubes can begin from the aromatic side chains, the replacement of an aromatic side chain with an aromatic side chain is favored (also chemical similarity explains this). The most basic facts about folding do not provide obvious support for the idea about flux tubes between residues.

i) Hydrophobic residues tend to cluster in dense packing in protein interior (antimatter at quark level) and Val ( $T$ ), Leu ( $T$ ), Ile ( $T$ ), Phe ( $T$ ), Ala ( $C$ ), and Gly ( $G$ ) make 63 percent of the interior of protein: the special role of Gly (matter rather than antimatter) is due to the reduction of the side chain to hydrogen atom.

ii) Asp ( $A$ ), Glu ( $A$ ), Lys ( $A$ ) and Arg ( $G$ ) with ionized residues are mostly at the surface of protein and make 23 per cent of protein surface and 4 per cent of interior. As noticed earlier, matter and antimatter at quark level tend to be far from each other.

iii) Polar groups tend to be paired by hydrogen bonds and oppositely charged groups tend to be near each other. Acidic Cys residues tend to be in positions where they can form S-S bonds. This cannot be explained by  $Y - Y_c$  pairing nor by the presence of bonds connecting residues in the proposed scenario. Aromatic residues tend to have favorable electrostatic interactions with each other and with S, O and amino groups.

## 2.5 4-D spin glass energy landscape and code of catalytic action

There is a proposal that protein folding corresponds to a motion in a fractal spin glass energy landscape in presence of external perturbations due to the presence of water and leading to the bottom of some deep valley [6]. In TGD framework 3-D spin glass landscape is replaced by 4-D one [I1]. The vacuum degeneracy of Kähler action implies 4-D spin glass energy landscape in the sense that quantum jump sequences lead to space-time sheets representing asymptotic self organization patterns depending only weakly on the initial conditions (with respect to subjective time measured as quantum jumps). Proteins would be like skilled musicians possessing a repertoire of motor activities represented by deep valleys in 4-D spin glass landscape. This picture generalizes to the functioning of living matter in various scales and the quantum dynamics of brain is a natural application giving also connection with p-adicity since ultrametric topology is naturally associated with the

space of valley bottoms. In the case of catalytic reactions a quantum jump changing Planck constant for some magnetic flux tubes connecting some living biomolecules (DNA, RNA, aminoacids, water(?),...) and changing the lengths of these flux tubes could be the basic mechanism leading from a given valley to a new one and the reduction of the genetic code to single nucleotide or di-nucleotide code would code this quantum jumps.

To sum up, this proposal for the folding code - or rather, the code of entire biocatalysis - is so beautiful that it deserves to be killed: this should be easy for a professional biochemist. If the hypothesis survives, it would provide a royal road to the understanding of the catalytic bio-chemistry.

### **3 A simple quantitative model for protein folding and catalyst action**

Levinthal paradox states that if protein folding is a process in which protein checks for all possible conformations, folding would take astrophysical time. Small single domain proteins with lengths below 100 residues however fold in single step in millisecond time scale and longest folding times are measured in days. This suggests that protein folding is in some sense guided dynamical process and flux tubes would be the natural guides.

It is possible to construct a simple quantitative model for protein folding and catalyst action assuming a long range interaction mediated by flux tubes between aminoacids for which middle nucleotides are conjugates. The model is consistent with quantum criticality, and the general vision about 4-D spin glass landscape. The extremals are not completely deterministic just as vacuum extremals of Kähler action and only absolute minimization of energy selects minima. The minimalistic interpretation is that absolute minimization of energy stabilizes various unstable patterns generated spontaneously by ordinary chemical interactions such as alpha helices and beta sheets.

The principle is flexible enough to carry out this purpose but also poses strong constraints on how these patterns integrate to higher level structures. The disappearance of a subset of flux tubes does not spoil the extremal property although it increases its non-determinism and makes folding less predictable and in the case of binding sites it reduces the selectivity of catalyst action. The interpretation would be in terms of molecular ageing. The density of flux tubes can be seen as analog for the resolution of quantum measurement which is in a fundamental role in quantum TGD, as well as a direct correlate for cognitive and sensory resolutions. The model extends to

a model of catalyst dynamics if one the relative motion of reactant molecules is slow in the time scale of folding dynamics so that adiabaticity assumption makes sense. In the following I often use the basic data which can be found from [4] without explicit reference.

### 3.1 The model

Let us assign potential energy to the flux tube connecting  $i$ :th and  $k(i)$ :th aminoacid and depending only on the distance  $r_{i,k(i)}$ . What comes in mind first is the potential energy of harmonic oscillator:

$$V(r) = \frac{kr^2}{2} . \quad (1)$$

$k > 0$  corresponds to harmonic oscillator. Also  $k < 0$  is possible in which case the distance between aminoacid and its conjugate tends to be maximized in equilibrium: this option turns out to be the more plausible one and conforms also with the notion of quantum criticality. Besides this there is the constraint that the distances between aminoacid and its follower are constant:  $r_{i+1,i} = R$  . Using Lagrange multipliers this gives rise to the action

$$L = -E = -\frac{k}{2} \sum_i r_{i,k(i)}^2 + \sum_i \lambda_i r_{i+1,i}^2 . \quad (2)$$

Energy is the negative of this action for static solutions. One could consider also adding kinetic term to this action to describe the dynamics of folding. This action is hoped to give only a qualitative view about folding and the ordinary chemical interactions should fix the details of the folding and select between different folding patterns. Several aminoacid chains could be present and have mutual long range interactions.

The extrema of this action satisfy

$$\frac{\partial L}{\partial r_i^k} = 0 , \quad i = 1, \dots, N . \quad (3)$$

This gives the conditions

$$\begin{aligned} \lambda_{i+1} \bar{r}_{i+1,i} - \lambda_{i-1} \bar{r}_{i,i-1} &= -k \bar{r}_{i,k(i)} , \\ r_{j+1,j} &= R . \end{aligned} \quad (4)$$

The geometric content of these conditions is that the vectors  $\bar{r}_{i,k(i)}$ ,  $\bar{r}_{i+1,i}$ , and  $\bar{r}_{i,i-1}$  are in the same plane. Thus long range interactions of aminoacids with their conjugates dictate the local folding of the aminoacid chain but extremum property alone does not say much about the lengths of the flux tubes.

Suppose that  $\bar{r}_i$ ,  $\bar{r}_{i,k(i)}$ ,  $\bar{r}_{i,i-1}$ ,  $\lambda_{i-1}$  are known. Can one solve  $\lambda_{i+1}$  and  $\bar{r}_{i+1,i}$ ? Since the vectors are in the same plane, the linear dependence does not fix the direction of  $\bar{r}_{i+1}$  in this plane but only the value of  $\lambda_i$  in this plane once  $\bar{r}_{i+1}$  is fixed or vice versa. Therefore the direction in the plane remains un-determined and equations of motion are not fully deterministic as far as extremals are considered. Absolute minimization however eliminates this non-determinism by maximizing the distances  $r_{i,i(k)}$  for  $k > 0$  option. The expressions for  $\lambda_i$  result from elementary linear algebra by introducing dual basis of non-orthogonal basis defined by  $\bar{r}_{i,k(i)}$  and  $\bar{r}_{i,i-1}$ .

$$\begin{aligned} \lambda_{i+1} &= -k\bar{e}_{i+1} \cdot \bar{r}_{i,k(i)} \quad , \quad \lambda_{i-1} = -k\bar{e}_{i-1} \cdot \bar{r}_{i,k(i)} \quad , \\ \bar{e}_{i+1} \cdot \bar{r}_{i+1,i} &= 1 \quad , \quad \bar{e}_{i+1} \cdot \bar{r}_{i,i-1} = 0 \quad , \\ \bar{e}_{i-1} \cdot \bar{r}_{i+1,i} &= 0 \quad , \quad \bar{e}_{i-1} \cdot \bar{r}_{i,i-1} = 1 \quad . \end{aligned} \tag{5}$$

The non-determinism does not make it easy to find absolute minimum since non-determinism corresponds to circle  $(S^1)^{2N}$  for aminoacid sequence with  $N$  flux tube pairings.

One can find extremals in the following manner. Divide the aminoacids into hydrophilic and hydrophobic ones (matter *resp.* antimatter at quark level) depending on whether  $Y \in \{A, G\}$  or  $Y \in \{T, C\}$ . Fix the positions of aminoacids with  $Y \in \{T, C\}$  and chose the positions of aminoacids  $Y \in \{A, G\}$  in such manner that they are consistent with the planarity condition. This gives an extremal but not an absolute minimum.

The consistency with chemical interactions could remove the non-determinism to a high degree and the most natural approach is perhaps to start from candidate configurations expressible as combinations of various unstable structures predicted by chemical interactions and see whether they are consistent with and stabilized by the flux tube dynamics. The strong resemblance with the dynamics defined by absolute minimization of Kähler action predicting spin glass degeneracy associated with vacuum extremals of Kähler action and removed by small deformations to non-vacuum extremals raises the hope that the model indeed catches something essential about the notions of 4-D spin glass degeneracy and quantum criticality.

### 3.2 Some general consequences

What can one conclude from the variational equations?

a) Absolute minimization of energy is very powerful selection principle and is expected to choose highly symmetric configurations such as  $\alpha$  helices,  $\beta$  sheets, and more complex structures. If combined with adiabaticity assumption it could also allow to understand the dynamics of binding between two proteins and protein and DNA/RNA.

b) The extremals of  $k > 0$  action are mirror images of  $k < 0$  action so that the energy minimum for  $k > 0$  is energy maximum for  $k < 0$ . If energy minimization is applied also the choice of  $Y - Y_c$  flux tubes, the connected aminoacids should be as near as possible which would favor long linear structures as opposed to an almost spherical tightly packed structure. In light of this  $k < 0$  option looks more realistic.

c)  $A, G$  surplus for mRNA implies that  $Y = A, G$  aminoacids dominate unless the asymmetry is concentrated on the third nucleotide. In presence of asymmetry all aminoacids cannot pair. If the aminoacid is not paired, it does not experience the long range force, and one has

$$\lambda_{i+1}\bar{r}_{i+1,i} - \lambda_{i-1}\bar{r}_{i,i-1} = 0 . \quad (6)$$

Situation becomes non-deterministic and the portions of the aminoacid chain for which the aminoacids do not have a pair behave like random coils. This is encouraging since this kind of portions are present in folded aminoacids.

d) The disappearance of some flux tubes does not destroy a given solution of the conditions but makes it increasingly non-deterministic. The interpretation as a degradation or ageing at molecular level conforms with the interpretation of braiding as a basic characteristic of life. An attractive interpretation of the density of flux tubes is as correlate for resolution for cognition and sensory perception and motor action as counterpart of measurement resolution which is fundamental notion of quantum TGD.

e) The model leaves open which aminoacids in the chain are paired. The pieces of aminoacid chain having long range interactions correspond to separate pieces of the chain or even different chains. Aminoacids at the surface of the protein and in the interior of the protein would thus be connected by flux tubes. Even more, the pairing could occur between connected pieces of the aminoacid chain. For  $k < 0$  option this would be very natural since long distance between hydrophobic and hydrophilic parts of chain would minimize the interaction energy. For this option  $r_{i,k(i)}$  would be as large as possible subject to the condition from fixed chain length. Also



the connection pattern could evolve during the folding so that the pattern minimizing the energy is approached.

f) The constraints on the primary structure of paired structures are too strong. The flux tubes connections need not however create structures from scratch but only stabilize structures- such as  $\alpha$  helices and  $\beta$  sheets- generated spontaneously by ordinary chemical interactions. This stabilization does not require complete pairing of aminoacids between two structures involved and thus constraints on the primary structure are much weaker. The minimization of the number of flux tubes could be the underlying principle.

g) As an example consider two pieces of chain connected by flux tubes separated by pieces which have no flux tube connections and thus behave non-deterministically. In this case a good candidate for energy minimum is a solution for which the two portions of the sequence are mirror images in the sense that the portions are obtained from each other by the transformation  $\lambda_{k(i)} = -\lambda_i$ . The presence of paired mirror image structures in proteins is suggestive. This kind of solutions can be glued together very flexibly by introducing non-bonded portions of the chain between them.

### 3.3 Model for helical structures

$\alpha$  helix, which is only one member of a rich family of helical structures possible for aminoacid chains, serves as the first test for the model. As a matter fact, the specific properties of  $\alpha$  helix are not relevant for the model discussed.

a)  $\alpha$  helix has nearly vertical NH—O= hydrogen bond between  $i$ :th and  $i - 4$ :th aminoacid. Also  $(i, i - 3)$  and  $(i, i - 5)$  bondings are possible. There are 3.6 residues per turn so that the basic structural unit has 5 turns and consists of 18 aminoacids. One residue corresponds to a vertical translation of 1.5 Angstrom. Chain contains single aminoacid per length of about 3.8 Angstrom and the angular separation of subsequent aminoacids is 100 degrees in the planar projection.

b)  $\alpha$  helices are not stable unless stabilized by secondary coiling: their lifetime is of order  $10^{-5} - 10^{-7}$  seconds. Hence it is not clear whether one should regard  $\alpha$  helix as a non-deterministic piece of a chain whose conformation is temporarily fixed by hydrogen bonding. If flux tubes are involved, the instability could be due to their temporary disappearance. Reconnection process comes also into mind but it does not imply non-deterministic behavior. If the number of flux tubes and their connectedness pattern are also subject to variation,  $\alpha$  helix identified a basically non-deterministic solution made deterministic by ordinary chemical interactions does not represent ab-

solute minimum of energy anymore.

Consider now a model of  $\alpha$  helix assuming that flux tubes are present.

a) The condition that  $\bar{r}_{i,k(i)}$  lies in the same plane as  $\bar{r}_{i+1,i}$  and  $\bar{r}_{i,i-1}$  fixes the situation to a high degree. The  $k(i)$ :th aminoacid would be in almost horizontal plane.

b)  $k(i) = i \pm 1$  rule is the local solution to the conditions and gives absolute minimum of energy for  $k > 0$  option. Alpha helix would consists of pairs alternating pairs of residue and its conjugate connected by a flux tube:  $(A_1 - A_{1,c}) - (A_2 - A_{2,c}) - \dots$ : this obviously is quite too restrictive condition for the primary structure. For a pair  $(i, k(i) = i + 1)$  with  $(\lambda_{i+1} = 1, \lambda_{i-1} = 0)$  the constraint says nothing non-trivial. Random coil would result given as a special case  $\alpha$  helix but for general primary structure the model fails. This supports  $k < 0$  option.

c) Second possibility is natural for  $k < 0$  option. Assume that aminoacid sequence is curved in such a manner that  $k(i)$  corresponds to aminoacid in some other part of the sequence, perhaps  $\alpha$  helix itself or second  $\alpha$  helix. The direction of the normal of the plane defined by  $\bar{r}_{i+1,i}$  and  $\bar{r}_{i,i-1}$  is nearly vertical and has 5-turn periodicity in the linear case so that the conjugate aminoacids are in nearly horizontal direction in second portion of chain or in second chain. Poly(Pro) and Poly(Gly) helices, which correspond to  $Y = C$  and  $Y = G$  respectively are exceptional in that they admit both trans and cis peptide bonds. In these cases stabilizing flux tubes must connect the helix to some other part of the aminoacid sequence. The spontaneous generation of unstable helices in sequences consisting of mere Pro or Gly would support the assumption that the role of flux tubes is to stabilize.

i) Perhaps the most realistic option utilizes the fact that hydrogen bonding is a spontaneous process induced by ordinary hydrogen bonding although free  $\alpha$  helix is not stable. Only some minimum number of flux tubes between  $\alpha$  helix and conjugate structure might be enough to guarantee stability. The condition that the vectors  $\bar{r}_{i,i+1}$  lie in the plane defined by  $\bar{r}_{i-1,i}$  and  $\bar{r}_{i,i(k)}$  for some values of  $i$  look rather weak but since the chain does not stretch this condition could be rather strong.

ii) At the other extreme would be complete coding. The relatively small deviation of the normal from vertical and its slow variation raises the hope that  $\bar{r}_{i,k(i)}$  could vary its direction sufficiently fast for  $k(i) = i + I$  rule to make  $\alpha$  helix possible or even the situation in which nearly antiparallel portions of  $\alpha$  helix code for each other. The distance between helices would be in this case highly relevant and cannot be too large. Double helices and hairpin like structures come in mind (hairpin like structures appear at the surface of protein where sudden changes of direction of chain take place).

DNA double helix could also involve similar dual coding. Locally  $\alpha$  helix structure would be due to local chemical interactions but stabilized by long range interactions.

iii) A more intricate possibility is scaling  $k(i) = n \times i + I$  in order to achieve a faster variation of  $\bar{r}_{i,k(i)}$ . There are two extreme situations. Assume two  $\alpha$  helices (possible connected by non-deterministic portion of chain) consisting of  $N$  aminoacids and assume  $k(i) = n \times i + I \text{ mod } N$ ,  $n \ll N$ . Effectively the  $\alpha$  helices would be circular and the coding would be analogous to the map  $\phi \rightarrow n\phi$  of circle replaced with regular N-polygon. The conjugacy property obviously poses strong additional conditions on the primary structures. For larger distances  $r_{i,k(i)}$  the map  $i \rightarrow k(i)$  can be however modified in some limits it might be possible to achieve consistency with the primary structure. The second extreme is obtained by picking from the non-deterministic part of the chain suitably the conjugate aminoacids connected to  $\alpha$  helix and allowing the rest of aminoacids to remain random. For this option the non-deterministically varying parts of the helix would involve some hidden determinism.

There are also more complex structures formed from helices. For coiled coils of two or more alpha helices consisting of repeating heptad unit of 7 aminoacids first and fifth aminoacids tend to be conjugates so that horizontal flux tubes connecting first and fifth aminoacids of neighboring could be responsible for the stability and make also possible the hydrophobic bonding between first and fourth residues. Collagen is a triplet helix and appears as a basic constituent of bones, tendons, skin, ligaments, blood vessels, and supporting membranous tissues. Collagen triple helix consists of very long repetitive sequences of type  $(Gly - X_{aa}Y_{aa})_n$ , with a preponderance of Pro for  $X_{aa}$  and  $Y_{aa}$  (also Lys residues are possible). Heating of collagen triple helix unfolds it and converts it to gelatin, in which polypeptide chains are dissociated, unraveled and disordered. Cooling regenerates these conformations for short stretches. That *Gly* ( $Y = G$ ) and *Pro* ( $Y = C$ ) are conjugate aminoacids supports the hypothesis about effective base-base pairing (Lys corresponds to  $Y = A$ ). The stabilizing mechanism would be nearly horizontal flux tubes between Gly and Pro residues. The mutations of Gly and Pro residues in the hydrophobic interior of the protein affect most the protein folding. This could be due to special geometric properties of Gly and Pro but could also mean that flux tubes between Gly and Pro tend are of special importance.

### 3.4 Model for $\beta$ sheets

$\beta$  sheets consist of typically 4-5 aminoacids long  $\beta$  strands which can be either parallel or antiparallel and are glued together by hydrogen bonds. Beta sheets are also slightly twisted which relates to the chirality of aminoacids. In antiparallel case strand returns back and forms at the ends of sheet a loop so that so called  $\beta$  hairpin is formed. In parallel case the strand returns as alpha helix to the lower end of the sheet.  $\beta$  strands have 2 aminoacids per turn so that  $\bar{r}_{i-1,i}$  and  $\bar{r}_{i,i+1}$  span a vertical plane. At the time of writing of [4] the mechanism of formation of  $\beta$  sheets was not understood.

The first thing to notice is that some minimal number of flux tubes should be enough to guarantee the stability of  $\beta$  sheet. Antiparallel  $\beta$  sheets appear always in pairs. In presence of twisting  $\bar{r}_{i-1,i}$  and  $\bar{r}_{i,i+1}$  span a plane very nearly orthogonal to the sheet. Therefore flux tubes connecting the  $\beta$  sheets horizontally could be responsible for the stabilization. In the case of twisted  $\beta$  sheet consisting of parallel strands the stabilizing flux tubes could connect  $\beta$  strand with parallel  $\alpha$  helix.

Secondary protein structures are divided into four classes on basis of their secondary structures. All these structures are consistent with the general model.

a) ( $\alpha$ ) containing only  $\alpha$  helices, which must stabilize each other by horizontal flux tubes.

b) ( $\beta$ ) containing only  $\beta$  sheets both usually antiparallel, which appear always in pairs packing against each other. Horizontal flux tubes connecting the  $\beta$  sheets must act as stabilizers.

c) ( $\alpha + \beta$ ) proteins can contain only single  $\beta$  sheet, usually antiparallel, with  $\alpha$  helices clustering together at one or both ends of the  $\beta$  sheet. Antiparallel  $\beta$  sheet stabilizes itself.

d) ( $\alpha/\beta$ ) in which sheets and helices interact and often alternate along the polypeptide chain. Single parallel  $\beta$  sheet and so called  $\beta$  barrel, kind of sandwich like structure, are basic examples here. The most spectacular barrel consists of 4+4 parallel  $\beta$  strands with  $\alpha$  helices outside the barrel.

To sum up, it seems that horizontal flux tubes might provide quite general stabilization mechanism and that parity breaking implying the twisting and preferred helicity is an essential element of this mechanism in the case of  $\beta$  sheets.

### 3.5 Model for protein-protein binding sites

Binding sites obey geometric complementarity and are known to resemble protein interior being closely packed. This is also taken to mean that aminoacid chains run parallel to the surface although this statement is not made explicitly in [4]: one could see binding sites as part of interior which is in a direct contact with exterior, somewhat like a sensory organ like eye. The interface between similar sized proteins is large and tends to be flat (not expected if proteins make sharp turns at the interface rather than running parallel to the surface). Various bonds eliminate electromagnetic interactions at the interface.

The question is whether complementary of bonded aminoacids should induce the geometric complementarity of the binding sites in the proposed model.

a) The binding sites could be connected by only very few flux tubes or flux tubes could connect all aminoacids in pairwise manner: the first extreme is highly flexible whereas second extreme would produce maximal selectivity. Complementary can thus be partial and its degree is predicted to correlate with selectivity. The interpretation of disappearance of flux tubes as molecular ageing conforms with the gradual loss of selectivity implying reduced performance of immune system.

b) The condition that the line connecting aminoacid in protein and its conjugate in the second protein lies in the plane spanned by  $\bar{r}_{i+1,i}$  and  $\bar{r}_{i,i-1}$  (and in corresponding complementary plane) requires that the aminoacid sequence at the surface is slightly curved in the direction of the conjugate aminoacid. From the example of [4] about the interface of identical proteins in the quaternary structure of dimer one learns that the geometrically and physically conjugate interfaces of identical monomers pair to form sandwich like structures via so called isologous and heterologous pairings such that valleys and hills fit. The interfaces are reported to resemble closely packed protein interiors and contain hydrophobic residues in the center and hydrophilic residues at periphery. Conjugacy requires that aminoacids in the periphery of the interface are connected to the central region of the conjugate interface and vice versa so that flux tubes cross each other and their lengths can be maximized even during contact. Since binding sites share all these properties one expects that the model applies also to them.

It would be nice to understand geometric and physical conjugation in terms of the genetic code. Aminoacid conjugation does not correspond in this model to the geometric conjugation of catalyst and its ligand but to the inversion mapping periphery and center (outer surface and interior) to

each other. Geometric and physical conjugation (acids and basics combine in the interface) means that a virtual protein A+B is cut to pieces along the surface in the interior defining the interfaces. Could this chopping of bigger proteins to smaller ones able to bind allow realization at the level of genome? And could also protein interior involve pairings analogous to catalyst and ligand pairings? This would partially explain why protein folding is more sensitive to the mutations in the interior of protein.

Flux tubes need not connect periphery of the interface of A/B to center of B/A all the time. Rather, they could be A-A and B-B type in inactive state and only during activated state the reconnection process would transform them to A-B and B-A type bonds and involve also a phase transition increasing Planck constant so that flux tubes get in reconnection contact at large enough distance. After that reduction of Planck constant would bring interfaces in contact. The center would be like pupil of eye to which the light rays from the periphery of second eye converge.

c) Slow enough relative motion of molecules induces an adiabatic variation of the shapes of the binding sites so that lock and key mechanism becomes dynamical. The simplest possibility is that binding site and its conjugate behave like two eyeballs gazing each other as proteins move with respect to each other. This is possible if binding sites are separated from the rest of the protein by nondeterministic pieces of chain. The analogy with eye might be actually deeper: I have proposed long time ago that directed attention in vision has as a space-time correlate flux tubes of topological light rays or both of these. Wormhole magnetic flux tubes might indeed connect perceiver and the object perceived and serve as correlates of attention in macroscopic length scales.

To sum up, the model of folding might be flexible enough to explain basic qualitative aspects of folding. The overall view seems to be that flux tubes force the portions of aminoacid sequences are forced to form representation about each other in their own geometry. What is also nice that the notions of finite measurement resolution and cognitive resolution which are fundamental notions of quantum TGD have direct correlates at the level of flux tube dynamics.

## 4 Some simple tests for the flux tube model of protein folding

I am grateful for Dale Trenary for suggesting a simple test for the model using data for the variants 3FIS and 1FIA of DNA binding protein Fis, and

1BHM representing an instance of type II restriction enzyme BamHI. All these are dimers of homonomers involving alpha helices. Dale has also provided the data about them from RCSB protein data bank [7, 8, 9]. The homonomers of 3FIS and 1FIA have 98 residues of which the position data for 73 are given in the data basis. For 1BHM the positions are known for 209 residues.

#### 4.1 The test

The test of the model is simple.

a) Denote by  $r_i$  the position vector of aminoacid  $i$  in given monomer. Calculate the normals  $n_i$  of the planes  $P_i$  associated with aminoacid  $i$  and defined by the vectors  $r_{i+1,i} = r_{i+1} - r_i$  and  $r_{i,i-1} = r_i - r_{i-1}$  of position vectors of  $i$ :th and  $i - 1$ :th amino-acid.  $n_i$  is given by the cross product of  $r_{i+1,i} \times r_{i,i-1}$  normalized to unity. The position of the = O atom can be taken to represent the position of aminoacid since flux tubes are assumed to connect = O atoms.

b) Count the number of amino-acid pairs  $(i, j)$  of type A-A, B-B or A-B for which the normals of the planes  $P(i)$  and  $P(j)$  form angle  $\theta < \theta_{max}$  or  $\pi - \theta < \theta_{max}$ . Denoting by  $n_{i,j}$  the unit vector in the direction of the vector  $r_i - r_j$ , this means that the conditions

$$|n_i \cdot n_{i,j}| < \cos(\theta_{max}) \quad , \quad |n_j \cdot n_{i,j}| < \cos(\pi/2 - \theta_{max}) = \sin(\theta_{max})$$

are satisfied. The set of allowed directions of normals corresponds to a band at the equator having angular weight  $2\theta_{max}$ .  $\sin(2\theta_{max}) = .34$  corresponds to the angle of 20 degrees which is average relative orientation for two alpha helices involving several alpha helices and can be taken as a reasonable first guess  $\sin(2\theta_{max})$  giving  $\theta_{max} = .17$  (10 degrees).

c) Count the number of conjugate aminoacid pairs of type A-A, B-B, and A-B where conjugacy of the pair means that middle nucleotides  $Y$  for the codons  $XYZ$  coding for the aminoacids in question are conjugate. In the case of ser this condition is not unique since ser is coded by codons in both second and fourth row of the code table so that a flux tube connecting two sers are possible. In the case considered the portions of monomer about which data are given contain however only single ser so that this does not affect the estimates. The sers of A and B could be however connected by a flux tube.

d) Count the numbers  $n_{A-A}$ ,  $n_{B-B}$ , and  $n_{A-B}$  of aminoacid pairs of type A-A, B-B, and A-B satisfying both the condition that aminoacids are

conjugates and can be connected by flux tube. It can happen that same aminoacid can be connected to several aminoacids and only one option can be chosen. This can be taken into account in the estimate. The relative positive vector connecting two subsequent aminoacids belongs automatically to the planes associated with these aminoacids and one must eliminate these pairs from the counting.

The following table summarizes the results of the calculation. The calculations are done for three proteins: for 1FIA and 3FIS and for 1BHM.

|                                             |           |           |           |           |
|---------------------------------------------|-----------|-----------|-----------|-----------|
| $\sin(\theta_{max}) = .1735$ (9.99 degrees) | $n_{A-A}$ | $n_{B-B}$ | $n_{A-B}$ | $N_{A-B}$ |
| 1BHM                                        | 54        | 54        | 94        | 286       |
| 1FIA                                        | 10        | 7         | 24        | 30        |
| 3FIS                                        | 4         | 8         | 17        | 21        |
| $\sin(\theta_{max}) = .05$ (2.86 degrees)   | $n_{A-A}$ | $n_{B-B}$ | $n_{A-B}$ | $N_{A-B}$ |
| 1BHM                                        | 24        | 21        | 17        | 18        |
| 1FIA                                        | 0         | 0         | 2         | 2         |
| 3FIS                                        | 0         | 0         | 2         | 2         |

Table 1.  $n_{A-A}$ ,  $n_{B-B}$  and  $n_{A-B}$  give the numbers of conjugate pairs of aminoacids that can be connected by flux tubes for two values of  $\theta_{max}$ . Only single flux tube is allowed to begin from a given aminoacid and the flux tubes connecting subsequent aminoacids are excluded.  $N_{A-B}$  gives the numbers of A-B type flux tubes without the first restriction. For a given value of  $\theta_{max}$  the angles between the normals of planes  $P(i)$  and  $P(j)$  and the relative position vector  $r_{i,j}$  satisfies  $\cos(\theta) < \sin(\theta_{max})$ . Numbers are given for  $\sin(\theta_{max}) = .1735$  (9.99 degrees) and for  $\sin(\theta_{max}) = .05$ .

## 4.2 About the interpretation of the results

Consider first X-X type flux tubes.

a) From Table 1 one finds that for  $\sin(\theta_{max}) = .05$  there are no aminoacid pairs of type A-A or B-B for 3FIS and 1FIA. The assumption that long flux tubes of type X-X are involved with the stabilization of alpha helices gives the lower bound  $\sin(\theta_{max}) > .11$  in the case of 3FIS.

b) The value  $\theta_{max} \simeq 10$  degrees relates to the twist angle  $\theta_{twist}$  of alpha helix defined as  $\tan(\theta_{twist}) = z/d$ , where  $z$  and  $d$  are vertical and horizontal distances between neighboring aminoacids. For  $d$  and  $R$  one obtains estimates by using the basic data about alpha helix (the radius is  $R = 2.7$  Angstroms, there is a vertical translation of  $z = 1.5$  Angstrom per aminoacid, and there are 3.6 aminoacids per turn giving  $\theta = 100$  degrees for the angle between subsequent aminoacids) as



$$\theta_{twist} = \arctan\left[\frac{z}{R\sqrt{(\cos\theta - 1)^2 + \sin^2\theta}}\right] \simeq 19.93 \text{ degrees}$$

giving  $\sin(\theta_{twist}) = .3409$  to be compared with  $\sin(2\theta_{max}) \simeq .34$  deduced from  $\theta_{max} \sim 10$  degrees. Therefore the planes  $P(i)$  associated with alpha helix could not be distinguished from horizontal ones as far as flux tubes are considered. The interactions responsible for the formation of alpha helices would sum up with the interaction induced by the flux tubes and modify the predictions of the simplest model.

Consider next A-B type flux tubes.

a)  $\sin(\theta_{max}) = .05$  represents a lower bound for the angular resolution if one requires that the flux tubes between two monomers stabilize the dimer in the case of 3FIS and 1FIA. As found, X-X type stabilization requires  $\sin(\theta_{max}) > .11$ .

b) In the case of  $N_{A-B}$  one can check what a model assuming random orientations for the planes  $P(i)$  predicts. Assume that the orientations of planes associated with aminoacids are random and that the numbers of aminoacids of type  $Y = A, G, T, C$  are roughly the same. The condition  $\cos(\theta) < \sin(\theta_{max})$  states that the allowed planes correspond to the directions of normals in the band of width  $2\theta_{max}$  at the equator of a sphere with the relative position vector  $r_{i,j}$  going through its poles. The probability of the normal vector to be inside this band equals to  $\sin(\theta_{max})$  in good approximation. The probability that aminoacid of type monomer X is connected to a given aminoacid of monomer Y by a flux tube is

$$p = \frac{1}{4} \times \sin^2(\theta_{max}) .$$

The factor 1/4 comes from the condition of conjugacy. This would give for the total number of flux tubes connecting aminoacid and conjugate without the restriction that only single flux tube begins from given aminoacid the estimate  $N_{A-B}(random) = p \times N(A)N(B)/2 \simeq N(A)N(B)\sin^2(\theta_{max})/8$ . The actual values of  $N_{A-B}(real)$  do not differ very much from those predicted by randomness hypothesis as Table 2 demonstrates. For  $\sin(\theta_{max}) = .1735$  one has  $N_{A-B}(random)/N_{A-B}(real) \in \{1.74, 1, 1.5, 1.05\}$  for 1BHM, 1FIA and 3FIS respectively. The three cases studied do not allow to decide whether flux tubes play a role in the protein stabilization.

|                                             |                 |                   |
|---------------------------------------------|-----------------|-------------------|
| $\sin(\theta_{max}) = .1735$ (9.99 degrees) | $N_{A-B}(real)$ | $N_{A-B}(random)$ |
| 1BHM                                        | 286             | 164               |
| 1FIA                                        | 30              | 20                |
| 3FIS                                        | 21              | 20                |
| $\sin(\theta_{max}) = .05$ (2.86 degrees)   | $N_{A-B}(real)$ | $N_{A-B}(random)$ |
| 1BHM                                        | 18              | 14                |
| 1FIA                                        | 2               | 1.7               |
| 3FIS                                        | 2               | 1.7               |

Table 2. Comparison of the predictions for  $N_{A-B}$  based on randomness and actual data  $N_{A-B}(real)$  for the three cases considered.

## 5 Appendix: Data about the monomers A and B of Fis protein

The following tables give the data relevant to the test of the protein folding in case of Fis protein.

|    |         |          |         |
|----|---------|----------|---------|
| -1 | 12.6610 | 16.6070  | 54.3020 |
| -2 | 11.8380 | 14.7110  | 51.5100 |
| 1  | 15.0390 | 13.6030  | 51.2950 |
| 2  | 14.4740 | 11.4460  | 54.0350 |
| 1  | 11.7900 | 9.6470   | 52.6430 |
| -2 | 14.0040 | 8.7830   | 50.1990 |
| 2  | 16.2610 | 7.3600   | 52.0040 |
| 2  | 14.2280 | 5.0800   | 53.4970 |
| -1 | 12.7610 | 3.8110   | 50.7830 |
| -2 | 16.1980 | 2.9040   | 49.8060 |
| 2  | 16.8050 | 1.0350   | 52.4120 |
| 2  | 13.6750 | -0.8500  | 52.3140 |
| 1  | 14.7210 | -1.6790  | 49.3000 |
| -2 | 18.0470 | -2.8280  | 49.4940 |
| -1 | 16.7930 | -5.2570  | 52.1510 |
| 2  | 14.1150 | -6.2710  | 48.6550 |
| -2 | 16.2930 | -7.9520  | 46.5050 |
| 2  | 19.7610 | -8.2010  | 45.5060 |
| 1  | 17.1530 | -10.0680 | 42.0540 |
| 2  | 17.2710 | -5.9940  | 42.3450 |

|    |          |         |         |
|----|----------|---------|---------|
| 2  | 14.3440  | -5.2090 | 39.1220 |
| -2 | 16.8600  | -2.0190 | 38.1360 |
| 2  | 14.0750  | 0.8040  | 34.9300 |
| 2  | 11.1900  | 3.1180  | 38.3290 |
| -2 | 12.1850  | 5.5390  | 40.5380 |
| 1  | 10.7460  | 8.0040  | 39.1580 |
| 2  | 7.4700   | 6.4710  | 38.7300 |
| -2 | 7.2540   | 5.8340  | 41.9590 |
| -2 | 7.6180   | 8.9980  | 42.9010 |
| -2 | 5.7000   | 10.6390 | 40.5630 |
| -1 | 2.9440   | 8.8350  | 41.8980 |
| 2  | 3.7800   | 10.0150 | 45.4970 |
| -2 | 3.7900   | 13.7890 | 43.5800 |
| 2  | 0.9210   | 13.9550 | 41.1160 |
| 2  | -1.3860  | 13.2010 | 43.1480 |
| -1 | -1.0880  | 15.5610 | 45.6060 |
| -2 | -0.7510  | 18.2200 | 43.5540 |
| -2 | -3.5090  | 17.9240 | 41.7620 |
| 2  | -5.5380  | 17.4780 | 44.7780 |
| -2 | -3.9360  | 20.5930 | 46.0260 |
| -2 | -5.1150  | 22.4320 | 43.0220 |
| -2 | -8.3660  | 21.2800 | 43.0770 |
| 2  | -8.9600  | 22.3790 | 46.1970 |
| 1  | -7.6230  | 25.5820 | 45.6200 |
| -1 | -9.4710  | 25.1240 | 41.5520 |
| 1  | -12.2730 | 23.1470 | 41.0330 |
| 1  | -10.3480 | 22.0380 | 37.5940 |
| 2  | -7.1890  | 24.4330 | 37.8550 |
| 2  | -4.8130  | 24.2160 | 34.9350 |
| -1 | -3.8380  | 27.3660 | 35.4160 |
| 1  | -2.8570  | 27.2540 | 38.6320 |
| -1 | -1.0330  | 24.5420 | 38.1990 |
| -1 | 0.9410   | 25.6280 | 35.8760 |
| -2 | 1.8300   | 28.4230 | 37.5480 |
| -2 | 2.7870   | 26.7820 | 40.6440 |
| -2 | 4.8580   | 24.5340 | 37.7740 |
| 1  | 5.6810   | 25.2810 | 34.2850 |
| -2 | 1.4860   | 24.9440 | 32.8550 |
| 2  | 0.2960   | 22.2960 | 31.1260 |
| 1  | -2.8820  | 20.9460 | 30.6430 |

|    |         |         |         |
|----|---------|---------|---------|
| 1  | -1.3030 | 18.4540 | 29.0750 |
| -1 | 0.4040  | 17.3870 | 31.6290 |
| -2 | -2.3360 | 17.0790 | 33.5970 |
| 1  | -3.6930 | 14.8620 | 31.1970 |
| 2  | -1.1100 | 12.8200 | 31.3080 |
| 2  | -1.3670 | 12.4820 | 34.8710 |
| -2 | -4.5120 | 11.4620 | 34.9350 |
| 2  | -4.1310 | 9.0300  | 32.8010 |
| 2  | -1.5020 | 7.3160  | 33.9630 |
| 1  | -4.0690 | 7.8060  | 37.4760 |
| 1  | -7.9040 | 8.6520  | 37.0160 |
| -2 | -7.1350 | 12.0380 | 35.3220 |
| 2  | -9.5870 | 16.1910 | 34.8440 |

Table 3. The data about monomer A. The integer  $n \in \{-2, -1, 2, 1\}$  in the first column tells whether the aminoacid corresponds to codon  $XYZ$ ,  $Y \in \{T, C, A, G\}$ . The remaining columns give Cartesian position coordinates of  $= O$  atom using Angström as a unit.

|    |         |         |         |
|----|---------|---------|---------|
| -1 | 4.4580  | 8.3420  | 52.8710 |
| -2 | 5.1580  | 10.7040 | 50.4990 |
| 1  | 1.8270  | 11.7430 | 49.8720 |
| 2  | 1.9780  | 12.6910 | 53.4490 |
| 1  | 4.4760  | 15.0640 | 52.5710 |
| -2 | 3.0500  | 16.6040 | 50.0980 |
| 2  | 0.5910  | 17.5260 | 51.8110 |
| 2  | 2.0530  | 19.0260 | 54.1820 |
| -1 | 3.9670  | 21.0300 | 52.1640 |
| -2 | 0.5390  | 22.2520 | 50.8760 |
| 2  | -0.5080 | 23.4430 | 53.6320 |
| 2  | 2.6860  | 24.9400 | 54.7800 |
| 1  | 2.0860  | 26.6310 | 51.9410 |
| -2 | -1.3600 | 27.8700 | 51.7600 |
| -1 | -0.6630 | 29.5330 | 54.4990 |
| 2  | 2.8560  | 30.8930 | 52.7840 |
| -2 | 0.5370  | 33.1190 | 51.2110 |
| 2  | -1.7420 | 35.2540 | 51.8130 |
| 1  | -2.5780 | 35.5550 | 47.8840 |
| 2  | -0.0350 | 32.7880 | 46.1280 |

|    |         |         |         |
|----|---------|---------|---------|
| 2  | 4.0250  | 32.6010 | 44.9800 |
| -2 | 2       | 30.3860 | 41.3180 |
| 2  | 6.0450  | 29.0120 | 38.3160 |
| 2  | 7.6640  | 25.2300 | 41.0910 |
| -2 | 6.3920  | 22.0990 | 42.5230 |
| 1  | 8.4530  | 20.2980 | 40.8240 |
| 2  | 11.1840 | 21.3930 | 41.5500 |
| -2 | 10.4240 | 21.4390 | 45.2600 |
| -2 | 10.3110 | 17.7780 | 44.4430 |
| -2 | 12.7100 | 16.7820 | 42.8400 |
| -1 | 15.3050 | 17.9480 | 44.7860 |
| 2  | 13.2230 | 16.1300 | 47.6190 |
| -2 | 14.1390 | 12.6910 | 44.7480 |
| 2  | 17.0870 | 13.1700 | 42.8120 |
| 2  | 19.4740 | 13.7440 | 45.5480 |
| -1 | 18.6070 | 10.5270 | 47.0040 |
| -2 | 18.9280 | 8.7090  | 44.3710 |
| -2 | 22.3020 | 9.7390  | 43.3670 |
| 2  | 23.2050 | 8.9910  | 46.6010 |
| -2 | 21.9140 | 5.5420  | 46.7090 |
| -2 | 23.2270 | 4.7900  | 43.5500 |
| -2 | 26.4710 | 5.4990  | 44.3250 |
| 2  | 26.4200 | 3.5380  | 46.8830 |
| 1  | 25.1400 | 0.7670  | 45.3070 |
| -1 | 28.1130 | 2.0490  | 42.1350 |
| 1  | 30.7420 | 3.7900  | 42.7790 |
| 1  | 29.7920 | 6.2190  | 39.8350 |
| 2  | 26.4100 | 4.0730  | 38.5660 |
| 2  | 24.7500 | 5.0180  | 35.9380 |
| -1 | 23.7400 | 2.2610  | 34.6970 |
| 1  | 22.3920 | 1.2520  | 37.6950 |
| -1 | 20.4710 | 4.2420  | 37.5880 |
| -1 | 19.0740 | 3.6600  | 34.7410 |
| -2 | 18.0900 | 0.5110  | 35.4880 |
| -2 | 16.4250 | 1.5060  | 38.6000 |
| -2 | 14.6950 | 4.6410  | 36.2110 |
| 1  | 15.2480 | 4.6320  | 32.0430 |
| -2 | 19.0180 | 5.3030  | 32.1130 |
| 2  | 20.7360 | 8.3170  | 31.5570 |
| 1  | 23.7170 | 9.9310  | 31.8650 |

|    |         |         |         |
|----|---------|---------|---------|
| 1  | 22.3490 | 13.0020 | 30.8540 |
| -1 | 20.3030 | 12.9600 | 33.3960 |
| -2 | 22.5520 | 12.7740 | 35.7780 |
| 1  | 24.0250 | 15.6360 | 34.3780 |
| 2  | 21.7000 | 17.4690 | 34.8570 |
| 2  | 20.8190 | 16.7430 | 38.1050 |
| -2 | 24.0090 | 17.3170 | 39.1140 |
| 2  | 24.3220 | 20.5160 | 37.6940 |
| 2  | 21.1210 | 21.7170 | 39.5130 |
| 1  | 23.4720 | 19.7210 | 42.9560 |
| 1  | 26.5170 | 19.3080 | 42.7250 |
| -2 | 28.0200 | 15.3010 | 41.5560 |
| 2  | 30.2300 | 17.5640 | 38.3350 |

Table 4. The data about monomer A. The notations are same as in Table 3.

## References

### Online books about TGD inspired theory of consciousness and quantum biology

- [TGDclass] M. Pitkänen (2006), *Physics in Many-Sheeted Space-Time*.  
<http://www.helsinki.fi/~matpitka/tgdclass/tgdclass.html>.
- [TGDconsc] M. Pitkänen (2006), *TGD Inspired Theory of Consciousness*.  
<http://www.helsinki.fi/~matpitka/tgdconsc/tgdconsc.html>.
- [TGDdeeg] M. Pitkänen (2006), *TGD and EEG*.  
<http://www.helsinki.fi/~matpitka/tgdeeg/tgdeeg/tgdeeg.html>.
- [TGDfree] M. Pitkänen (2006), *TGD and Fringe Physics*.  
<http://www.helsinki.fi/~matpitka/freenergy/freenergy.html>.
- [TGDgeme] M. Pitkänen (2006), *Mathematical Aspects of Consciousness Theory*.  
<http://www.helsinki.fi/~matpitka/genememe/genememe.html>.
- [TGDgeom] M. Pitkänen (2006), *Quantum Physics as Infinite-Dimensional Geometry*.  
<http://www.helsinki.fi/~matpitka/tgdgeom/tgdgeom.html>.

- [TGDholo] M. Pitkänen (2006), *Bio-Systems as Conscious Holograms*.  
<http://www.helsinki.fi/~matpitka/hologram/hologram.html>.
- [TGDmagn] M. Pitkänen (2006), *Magnetospheric Consciousness*.  
<http://www.helsinki.fi/~matpitka/magnconsc/magnconsc.html>.
- [TGDmathc] M. Pitkänen (2006), *Mathematical Aspects of Consciousness Theory*.  
<http://www.helsinki.fi/~matpitka/magnconsc/mathconsc.html>.
- [TGDnumber] M. Pitkänen (2006), *TGD as a Generalized Number Theory*.  
<http://www.helsinki.fi/~matpitka/tgdnumber/tgdnumber.html>.
- [TGDpad] M. Pitkänen (2006), *p-Adic length Scale Hypothesis and Dark Matter Hierarchy*.  
<http://www.helsinki.fi/~matpitka/paddark/paddark.html>.
- [TGDquant] M. Pitkänen (2006), *Quantum TGD*.  
<http://www.helsinki.fi/~matpitka/tgdquant/tgdquant.html>.
- [TGDselforg] M. Pitkänen (2006), *Bio-Systems as Self-Organizing Quantum Systems*.  
<http://www.helsinki.fi/~matpitka/bioselforg/bioselforg.html>.
- [TGDview] M. Pitkänen (2006), *Topological Geometrodynamics: Overview*.  
<http://www.helsinki.fi/~matpitka/tgdview/tgdview.html>.
- [TGDware] M. Pitkänen (2006), *Quantum Hardware of Living Matter*.  
<http://www.helsinki.fi/~matpitka/bioware/bioware.html>.

## References to the chapters of books

- [H2] The chapter *Negentropy Maximization Principle* of [TGDconsc].  
<http://www.helsinki.fi/~matpitka/tgdconsc/tgdconsc.html#nmpe>.
- [H11] The chapter *TGD Based Model for OBEs* of [TGDconsc].  
<http://www.helsinki.fi/~matpitka/tgdconsc/tgdconsc.html#OBE>.
- [I1] The chapter *Quantum Theory of Self-Organization* of [TGDselforg].  
<http://www.helsinki.fi/~matpitka/bioselforg/bioselforg.html#selforgac>.
- [J1] The chapter *Bio-Systems as Super-Conductors: part I* of [TGDware].  
<http://www.helsinki.fi/~matpitka/bioware/bioware.html#superc1>.

- [J5] The chapter *Wormhole Magnetic Fields* of [TGDware].  
<http://www.helsinki.fi/~matpitka/bioware/bioware.html#wormc>.
- [L5] The chapter *DNA as Topological Quantum Computer* of [TGDgame].  
<http://www.helsinki.fi/~matpitka/genememe/genememe.html#dnatqc>.
- [L6] The chapter *Pre-Biotic Evolution in Many-Sheeted Space-Time* of [TGDgame].  
<http://www.helsinki.fi/~matpitka/genememe/genememe.html#prebio>.
- [M2] The chapter *Quantum Model for Nerve Pulse* of [TGDeeg].  
<http://www.helsinki.fi/~matpitka/tgdeeg/tgdeeg/tgdeeg.html#pulse>.

## Other references

- [1] S. Rackovsky (1993), *On the nature of the protein folding code*, Proc. Natl. Acad. Sci. USA. January 15; 90(2): 644648.  
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=45720>.  
*Protein folding*, [http://en.wikipedia.org/wiki/Protein\\_folding](http://en.wikipedia.org/wiki/Protein_folding).
- [2] *Alpha helix*, [http://en.wikipedia.org/wiki/Alpha\\_helix](http://en.wikipedia.org/wiki/Alpha_helix).  
*Beta sheet*, [http://en.wikipedia.org/wiki/Beta\\_sheet](http://en.wikipedia.org/wiki/Beta_sheet).
- [3] *List of standard aminoacids*, [http://en.wikipedia.org/wiki/List\\_of\\_standard\\_amino\\_acids](http://en.wikipedia.org/wiki/List_of_standard_amino_acids).
- [4] T. E. Creighton (1993), *Proteins: Structures and Molecular Properties*, W. H. Freeman and Company. New York.
- [5] *Phospholipids*, <http://en.wikipedia.org/wiki/Phospholipids>.
- [6] M. Buchanan (1998), *Shape is all*, New Scientist No 2157.
- [7] H.S.Yuan, S.E.Finkel, J.A.Feng,M.Kaczor-Grzeskowiak, R.C.Johnson, R.E.Dickerson (1991), *The molecular structure of wild-type and a mutant Fis protein: relationship between mutational changes and recombinational enhancer function or dna binding*, Proc.Natl.Acad.Sci.Usa, vol. 88, 9558. The URL of the protein data bank is <http://www.rcsb.org/pdb/home/>.  
*DNA binding protein 3fis*, RCSB Protein Data Bank, <http://www.ignmtest.cccb.pitt.edu/cgi-bin/anm/anm1.cgi>.  
[http://en.wikipedia.org/wiki/DNA-binding\\_protein](http://en.wikipedia.org/wiki/DNA-binding_protein).



- [8] Kostrewa, D., Granzin, J., Stock, D., Choe, H.W., Labahn, J., Saenger, W. (1992), *Crystal structure of the factor for inversion stimulation FIS at 2.0 Å resolution*. J. Mol. Biol. 226: 209-226.
- [9] *BamHI*, <http://en.wikipedia.org/wiki/BamHI>.  
Newman, M., Strzelecka, T., Dorner, L.F., Schildkraut, I., Aggarwal, A.K. (1995), *Structure of Bam HI endonuclease bound to DNA: partial folding and unfolding on DNA binding*. Science 269: 656-663.

1fia:

Dale