

Design of copolymeric materials

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

1995 J. Phys. A: Math. Gen. 28 L387

(<http://iopscience.iop.org/0305-4470/28/14/003>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 171.66.16.68

The article was downloaded on 01/06/2010 at 23:52

Please note that [terms and conditions apply](#).

LETTER TO THE EDITOR

Design of copolymeric materials

Tanya Kurosky and J M Deutsch

Physics Department, University of California, Santa Cruz, CA 95064, USA

Received 31 May 1995

Abstract. We devise a method for designing materials that will have some desired structural characteristics. We apply it to multiblock copolymers that have two different types of monomer, A and B. We show how to determine what sequence of A's and B's should be synthesized in order to give a particular structure and morphology. Using this method in conjunction with the theory of microphase separation developed by Leibler, we show it is possible to efficiently search for a desired morphology. The method is quite general and can be extended to design isolated heteropolymers, such as proteins, with desired structural characteristics. We show that by making certain approximations to the exact algorithm, a method recently proposed by Shakhnovich and Gutin is obtained. The problems with this method are discussed and we propose an improved approximate algorithm that is computationally efficient.

The problem addressed in this letter is the following. Is there an efficient method for designing a material with a particular morphology or structure? We develop a systematic approach to this problem that we illustrate for the design of copolymeric materials.

Structures of single chains in solution have also been extensively studied, often in relation to the important biological question of how to determine the structure of a protein from its sequence [1, 2]. Work on the design of a chemical sequence which has a desired three-dimensional structure has also been recently considered [3]. For the two-dimensional model of Dill *et al* [2] it has been possible to devise a set of rules that achieve the desired tertiary structure [4]. In three dimensions, much less is known. *Ad hoc* methods have been attempted [5, 6] but recent tests have shown that they are not entirely efficacious [7].

To illustrate our general method, we will apply it to design of block copolymers, which are polymers made out of more than one chemical species. We will consider copolymeric systems made up of two constituent types of monomer denoted A and B. The phase diagram of such materials has been studied as a function of the Flory interaction parameter χ and the lengths of the segments of A and B. Lamellar, hexagonally closed packed, body centred cubic [8], and gyroid [9] phases have been predicted. Experimentally copolymers have been found to exhibit a variety of different structures, sometimes referred to as microphases.

Suppose we would like to design a new phase that has a given symmetry. Until now it was necessary to do an inefficient search through the phase diagram of the system in order to find the desired symmetry. We show below that we can find a function that when minimized can home in on the correct structure. We then implement this practically in the framework of the theory of copolymers developed by Leibler [8]. We then turn to the problem of protein design and discuss how the method we have developed can be used in this context.

First we wish to determine the correct function to minimize in order to obtain the best sequence corresponding to the desired morphology. We begin with a formulation of the general problem that we wish to solve. Consider a system with coordinates denoted by

Γ and with a chemical sequence denoted by S . We can define a function which tells us whether each structure Γ is in the desired set of structures. Call this $P_{\text{struct}}(\Gamma)$. It is a constant if Γ belongs to the class of desired structures and 0 otherwise. In practice we will express $P_{\text{struct}}(\Gamma)$ in terms of the clamping potential defined below. Consider next the probability that a sequence S gives a desired structure. Since this is a conditional probability we denote it by $P(S|\text{struct})$. We wish to find the maximum of $P(S|\text{struct})$ over all sequences S keeping the structure fixed. To uniquely define $P(S|\text{struct})$, we have to choose the *a priori* probability of choosing an arbitrary sequence S . The simplest choice is that it is uniform, that is $P(S)$ is constant. We consider the system in equilibrium at a finite temperature so the probability that a sequence S has structure Γ is given by the Boltzmann factor $P(\Gamma|S) = \exp(-[H_S(\Gamma) - F_0(S)]/T)$, where $H_S(\Gamma)$ is the Hamiltonian of the system with a particular sequence S , and $F_0(S)$ is the corresponding free energy of the system with S kept constant. Bayes' theorem states that the joint probability $P(\Gamma, S) = P(\Gamma|S)P(S) = P(S|\Gamma)P(\Gamma)$, hence after some algebra, we obtain

$$P(S|\text{struct}) = \sum_{\Gamma} P_{\text{struct}}(\Gamma)P(S|\Gamma) = \sum_{\Gamma} P'_{\text{struct}}(\Gamma) \exp(-(H_S(\Gamma) - F_0(S))/T) \quad (1)$$

where

$$P'_{\text{struct}}(\Gamma) = \frac{P_{\text{struct}}(\Gamma)}{\sum_S \exp(-(H_S(\Gamma) - F_0(S))/T)}. \quad (2)$$

Because P'_{struct} does not depend on the sequence S , we can equally well regard it rather than P_{struct} as given. It can be thought of as imposing an external clamping potential on the system, pushing it into the correct structure, $P'_{\text{struct}}(\Gamma) \equiv \exp(-V_{\text{ext}}(\Gamma)/T)$. We will make an appropriate choice for the clamping potential, $V_{\text{ext}}(\Gamma)$ for each problem, based on physical reasoning rather than using (2) directly. Thus (1) can be further reduced to

$$P(S|\text{struct}) \propto \exp(-(F_{\text{struct}}(S) - F_0(S))/T) \equiv \exp(-\Delta F/T) \quad (3)$$

where $F_{\text{struct}}(S)$ is the free energy *pushed* into a certain structure by the clamping potential

$$\exp(-F_{\text{struct}}(S)/T) \equiv \sum_{\Gamma} \exp((-H_S(\Gamma) + V_{\text{ext}}(\Gamma))/T). \quad (4)$$

The physical interpretation of this is clear. The optimum sequence is the one that minimizes the difference ΔF between the unrestricted free energy and the free energy clamped in the desired structure. Intuitively this is reasonable because it picks out a sequence that naturally wants to spend a lot of time in this structure. This result is a generalization of that used in determining couplings in Boltzmann machines [10].

As a first example, we apply the formalism above to design copolymeric systems with desired lattice structures. Leibler [8] has developed a theory of microphase separation for di-block copolymers. Here we have further generalized this to m blocks each block being made up entirely of A monomers or of B monomers, the i th having a length l_i , $i = 1 \dots m$. The fractional length of the i th block is $f_i \equiv l_i/L$ where L is the total chain length. The A and B monomers have an incompatibility, or Flory, parameter χ giving the degree to which the A and B monomers wish to segregate. Leibler took as the order parameter $\Delta\rho$, the ensemble average of the difference between the density of A's and the average density of A's. He was able to construct an expansion of the free energy in terms of $\Delta\rho$ and calculated explicitly the expansion to fourth order in terms of the underlying chemical structure of the chains, that is, l_1 , l_2 , and χ . This theory should work well near the spinodal point for this system, because $\Delta\rho$ is small there.

To determine the stability of density variations with different crystallographic symmetries, Leibler took $\Delta\rho$ to be a periodic function of position \mathbf{r} ,

$$\Delta\rho(\mathbf{r}) = \sum_{j=1}^n \psi(\mathbf{q}_j) \exp(i\mathbf{q}_j \cdot \mathbf{r}) \quad (5)$$

choosing the \mathbf{q}_j 's, $j = 1, \dots, n$ to be the smallest non-zero reciprocal lattice vectors of the lattice structure being considered. The magnitude, q^* , of the \mathbf{q}_j 's is taken to be the wavevector at the spinodal point where divergent fluctuations first appear. He further took the magnitude, but not the phase, of all the $\psi(\mathbf{q}_j)$'s to be equal. By choosing different \mathbf{q}_j 's and minimizing the resultant free energies he computed which crystal structure had the lowest free energy. He was able to obtain a phase diagram for the system as a function of f_1 and χN . Besides the high-temperature disordered phase he found that the body centred cubic, triangular (hex.), and lamellar (lam.) phases existed in different regions of the phase diagram. Further work [11, 12] extended this treatment to triblocks.

It is convenient to take $V_{\text{ext}}(\Gamma)$ to be smoothly varying:

$$V_{\text{ext}}(\mathbf{r}) = -v \sum_{j=1}^n \exp(i\mathbf{q}_j \cdot \mathbf{r}). \quad (6)$$

The magnitude of v adjusts the degree to which $\Delta\rho$ fluctuates. A clamping potential with, for example, hexagonal symmetry is the sum of three plane waves. Applying such a potential to the unclamped free energy F_0 , we obtain a clamped free energy

$$F(V_{\text{ext}}) = F_0 + \int \rho(\mathbf{r}) V_{\text{ext}}(\mathbf{r}) d^3r \quad (7)$$

which will tend to push the system into a phase with the symmetry of the external potential.

Unfortunately in the general formalism developed above it is necessary for the clamping potential V_{ext} to be very large when monomers stray from the desired structure. This requirement leads to a large $\Delta\rho$ and hence is incompatible with the limits of validity of the free energy expansion of Leibler which is only valid in the limit of weak segregation. Therefore we need to consider complications that arise when only a weak clamping potential is applied.

For small V_{ext} , the system will not always be pushed into the symmetry of the external potential. In fact, for the values of v that we use, the effect of adding V_{ext} is only to slightly enlarge the region of the phase diagram that has the symmetry of V_{ext} . Our algorithm determines which phase the system is in by allowing the magnitudes of the $\psi(\mathbf{q}_j)$'s to be unequal and then minimizing with respect to them. This allows the possibility of mixed phases with more than one type of symmetry. For example if we consider the hexagonal phase and write (5) in terms of its wavevectors \mathbf{q}_1 , \mathbf{q}_2 , and \mathbf{q}_3 ,

$$\Delta\rho(\mathbf{r}) = \psi_{\text{lam}} \exp(i\mathbf{q}_1 \cdot \mathbf{r}) + \psi_{\text{hex}} (\exp(i\mathbf{q}_2 \cdot \mathbf{r}) + \exp(i\mathbf{q}_3 \cdot \mathbf{r})) \quad (8)$$

then for $\psi_{\text{lam}} = \psi_{\text{hex}}$ we have a hexagonal structure ($n = 3$ for the hex. phase), and for $\psi_{\text{hex}} = 0$ we have a lamellar structure ($n = 1$ for the lam. phase).

For small V_{ext} ,

$$\Delta F \equiv F(V_{\text{ext}}) - F_0 = v(\partial \Delta F / \partial v) = -v \sum_{j=1}^{n_c} \psi(\mathbf{q}_j) \quad (9)$$

where n_c is the number of \mathbf{q}_j 's that V_{ext} has in common with $\Delta\rho(\mathbf{r})$. From this it can be seen that ΔF does not do quite what one would like. In general a clamping potential will have reciprocal lattice vectors in common with more than one type of symmetry in

$\Delta\rho(r)$. This means that ΔF will be lowered for other phases in addition to the one with the symmetry of V_{ext} . For example, for a clamping potential with hexagonal symmetry V_{hex} , $\Delta F(V_{\text{hex}}) = F(V_{\text{hex}}) - F_0$ will be lowered for the lamellar phase as well as the hexagonal. We would like a functional which is more selective in order to design a hexagonal material. The solution in this case is to subtract the lamellar component as follows. Looking at (8), we see that the appropriate functional to maximize is ψ_{hex} . We would like to express ψ_{hex} in terms of free energy differences. From (9) $\psi_{\text{hex}} = -(\partial\Delta F(V_{\text{hex}})/\partial v) - \partial\Delta F(V_{\text{lam}})/\partial v)/2$. For small v this is proportional to $F(V_{\text{hex}}) - F(V_{\text{lam}})$. Therefore minimizing the difference in free energy between hexagonal and lamellar clamping will bring the system into the hexagonal phase. Figure 1(a) shows $(F(V_{\text{hex}}) - F(V_{\text{lam}}))$ as a function of the fraction f for diblock copolymers at $\chi = 20$. The minimum $f = 0.32$ occurs in the correct position indicated by Leibler's theory.

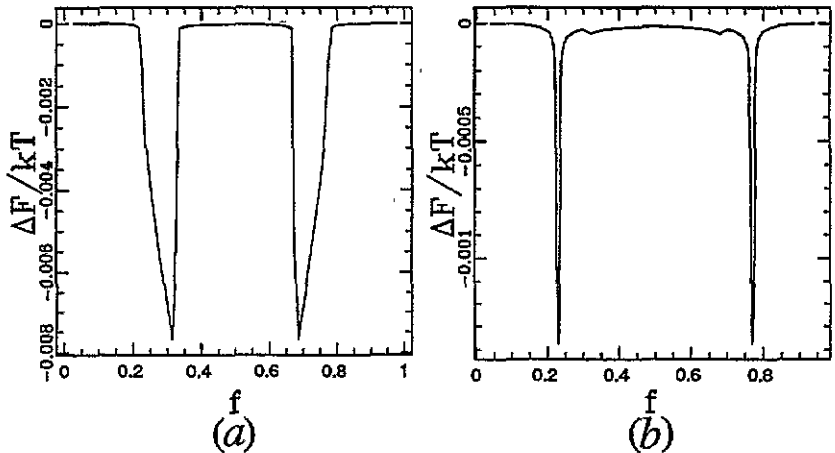


Figure 1. (a) $F(V_{\text{hex}}) - F(V_{\text{lam}})$ and (b) $F(V_{\text{BCC}}) - F(V_{\text{hex}})$, as a function of the fraction f for diblock copolymer design. The minima give the best choice for the design of hexagonal material and BCC material, respectively.

We also tested out this method for designing diblock BCC structures. Figure 1(b) shows $F(V_{\text{BCC}}) - F(V_{\text{hex}})$ as a function of f at $\chi = 20$. The minimum $f = 0.23$ also occurs in the correct position but there is also a secondary small minimum at $f = 0.32$. This shows that there is no guarantee that there will only be one minimum for this minimization function.

Now we turn to the problem of protein design. An interesting approach has recently been proposed by Shakhnovich and Gutin (SG) [5, 6], who proposed a method for solving this problem for a simplified lattice model using a self-avoiding chain. The model they employed has sequences $\{\sigma_i\}$ of two possible monomer types that are given values ± 1 , for chains of length N . This, plus the positions of all the monomers $\{r_i\}$, completely describe the state of the chain. We wish to find a sequence that causes the chain to fold up into a desired structure, but we do not care about what the monomer types end up being. The energy is

$$E(\{\sigma_i\}, \{r_i\}) = \frac{1}{2} \sum_{i,j}^N (B_0 + B\sigma_i\sigma_j)\Delta(r_i - r_j). \quad (10)$$

Here we will consider the case where $\Delta(r_i - r_j) = 1$ if r_i and r_j are nearest neighbours,

and is zero otherwise. One sets $B < 0$, since this favours ferromagnetic ordering of the σ 's, which means that the monomers will want to segregate, and $B_0 < 0$ since this provides an attractive interaction between monomers causing the protein to collapse.

Their method of sequence determination was to do a constrained minimization of the energy of the chain in sequence space. The constraint was that the total magnetization was held constant, in practice, close to zero. Note that without this constraint, all the σ 's would become equal, and one would have a homopolymer which does not have a well-defined structure. Unfortunately it appears that even with constrained minimization, sequences found do not necessarily have to have the desired structures as the lowest energy states [7]. Even if the desired structure is a ground state, there may be a large ground-state degeneracy in which case the structure is ill-defined, as in the case of the unconstrained minimization just mentioned.

The correct functional to minimize is ΔF defined in (3). If we specialize to the problem considered here, the constraining potential is a delta function since our structure is to be precisely determined. Calling the coordinates of the desired structure $\{r_i^0\}$, the correct functional to minimize is, according to (3),

$$\Delta F = E(\{\sigma_i\}, \{r_i^0\}) - F_0(\{\sigma_i\}). \quad (11)$$

SG's method is a minimization of only the first term, with a constraint of constant total magnetization. We will now argue that the second term is not negligible and cannot be omitted. In fact, we will see that a crude approximation gives an answer similar to the constraint of constant magnetization. But this approximation is of dubious validity which is why it failed.

Take B_0 to be very large and negative so that all stable structures must be globular with minimal surface area. The coordinates of our structure $\{r_i^0\}$ are therefore constrained to be of this compact type and the term in the energy involving B_0 will not vary and can now be ignored in the minimization. The space of all conformations $\{r_i\}$ we need to consider are also compact conformations of the same overall shape, but different internal arrangements. We now expand out $F(\{\sigma_i\})$ keeping only the lowest order cumulant. $F(\{\sigma_i\}) \approx \langle E \rangle + \text{constant}$. The angled brackets denote an average that is equally weighted over all compact conformations with minimal surface area. Ignoring constant terms, this gives $F_0(\{\sigma_i\}) \approx \frac{1}{2} B \sum_{i,j}^N \sigma_i \sigma_j \langle \Delta(r_i - r_j) \rangle$ so

$$\Delta F \approx \frac{1}{2} B \sum_{i,j}^N [\Delta(r_i - r_j) - \langle \Delta(r_i - r_j) \rangle] \sigma_i \sigma_j. \quad (12)$$

One can easily show that the nearest-neighbour interactions along the backbone of the chain cancel, because the probability that monomers i and $i + 1$ are next to each other is unity. This must happen because this kind of interaction is the same for all configurations and consequently cannot play a role in choosing the optimum σ 's. We now find an approximate functional form for $\langle \Delta(r_i - r_j) \rangle$. Since the chain is compact there is a short screening length. We therefore expect random walk correlations when $|i - j|$ is greater than a few lattice spacings. However, when $|i - j|^{1/2}$ becomes of order of the diameter of the protein, the conformations cease to look like random walks as the protein is compact. This crossover corresponds to $|i - j| \sim N^{2/3}$. For scales larger than this the correlation function should be almost constant. Therefore

$$\langle \Delta(r_i - r_j) \rangle \sim \begin{cases} |i - j|^{-3/2} & \text{for } 1 \ll |i - j| \ll N^{2/3} \\ 1/N & \text{for } |i - j| \gg N^{2/3}. \end{cases} \quad (13)$$

Therefore in this approximation, $F(\{\sigma_i\})$ looks like a one-dimensional Ising model with the above long-range interaction.

If we ignore the variation of $\langle \Delta(r_i - r_j) \rangle$ with $|i - j|$, i.e. $\langle \Delta(r_i - r_j) \rangle = 1/N$ so $\Delta F_{\text{MF}} = E(\{\sigma_i\}, \{r_i^0\}) - B(\sum_i^N \sigma_i)^2 / (2N)$, then F_0 gives an infinite-range mean-field contribution that is antiferromagnetic, and hence acts as a 'soft' constraint favouring a total magnetization of zero. To have a total magnetization of zero, one must introduce a domain wall, which increases $E(\{\sigma_i\}, \{r_i^0\})$ by order $N^{2/3}$, but this is more than compensated by the gain in free energy of the second term which is of order N . Hence in this limit we recover the approximation of SG. However, this is a rather drastic approximation. Even within the first-order expansion derived above, one is *not* justified in neglecting the shorter distance variation of $\langle \Delta(r_i - r_j) \rangle$ because, within this mean-field approximation, there is a spurious degeneracy in the σ 's that minimize ΔF_{MF} . This is because the three-dimensional arrangement of the σ_i 's namely $\sigma(r)$ is *independent* of the desired conformation $\{r_i^0\}$. That is, to find the correct sequence, it is *not* necessary to consider the different internal arrangements of the chain inside the compact cluster, as they all give an identical ΔF_{MF} . After the minimization of ΔF_{MF} has been performed once, the three-dimensional arrangement of the σ 's do not change when the desired conformation is changed.

However, a non-constant $\langle \Delta(r_i - r_j) \rangle$ breaks this degeneracy. It adds antiferromagnetic couplings along the backbone of the chain. This means that the domain wall should tend to orient itself roughly perpendicular to the direction of the backbone of the chain to satisfy the antiferromagnetic couplings. The contribution to ΔF due to the $|i - j|^{-3/2}$ decay of $\langle \Delta(r_i - r_j) \rangle$ is substantial and cannot be neglected. We can easily estimate it for the case of a desired structure $\{r_i^0\}$ that is in a typical random configuration. For a piece of arclength $s \equiv |i - j| = N^{2/3}$ the contribution to ΔF is of order $Bs^2/s^{3/2} = BL^{1/3}$. But there are $L/L^{2/3}$ such pieces, giving a total contribution of order $BL^{2/3}$. This is precisely the same order as the energy of the domain wall and therefore must be considered when designing proteins.

In conclusion, we have developed a method to design molecules that will self-assemble into a desired structure. We used this to design block copolymers with desired structural characteristics, within the framework of Leibler's mean-field theory.

We also note that the method for design described above should also work for the problem of protein design for which no trustworthy methods have so far been devised. It would be interesting to see how well the approximate minimization function, given by (12), designs stable proteins.

One of us (TK) wishes to thank Mark Dinan, Hemant Bokil and Doug Williams for useful discussions. We would also like to thank AP Young for useful comments and a critical reading of the manuscript. This work is supported by NSF grant DMR-9419362 and acknowledgment is made to the Donors of the Petroleum Research Fund, administered by the American Chemical Society for partial support of this research.

References

- [1] Creighton T E 1993 *Proteins: Structures and Molecular Properties* 2nd edn (New York: WH Freeman)
- [2] Dill K A 1990 *Biochemistry* **29** 7133
- [3] Chan H S and Dill K A 1991 *J. Chem. Phys.* **95** 3775
- [4] Yue K and Dill K A 1992 *Proc. Natl Acad. Sci. (USA)* **89** 4163
- [5] Shakhnovich E I and Gutin A M 1993 *Protein Eng.* **6** 793
- [6] Shakhnovich E I and Gutin A M 1993 *Proc. Natl Acad. Sci. (USA)* **90** 7195-6 793

- [7] Yue K, Fiebig K M, Thomas P D, Chan H S, Shakhnovich E I and Dill K A 1995 *Proc. Natl Acad. Sci. (USA)* **92** 325
- [8] Leibler L 1980 *Macromol.* **13** 1602
- [9] Matsen M W and Schick M 1994 *Phys. Rev. Lett.* **72** 2660
- [10] Hertz J, Krogh A and Palmer R G 1991 *Introduction to the Theory of Neural Computation* (New York: Addison-Wesley)
- [11] Mayes A M and Olvera de la Cruz M 1989 *J. Chem. Phys.* **91** 7228
- [12] Olvera de la Cruz M 1989 *J. Chem. Phys.* **90** 1995