
Integer Quadratic Programming Models in Computational Biology

Harvey J. Greenberg

Department of Mathematical Sciences, University of Colorado at Denver and Health Sciences Center, U.S.A.

`harvey.greenberg@cudenver.edu`

1 Introduction

This presentation has two purposes: (1) show operations researchers how they can apply quadratic binary programming to current problems in molecular biology, and (2) show formulations of some combinatorial optimization problems as integer programs. The former purpose is primary, and I wish to persuade researchers to enter this exciting frontier. The latter purpose is part of a work in progress.

I begin with some background, particularly in the biology; mathematical programming terms can be found in the *Mathematical Programming Glossary* [10]. I then present four integer quadratic programming (IQP) models. My last section indicates some avenues for research.

2 Background

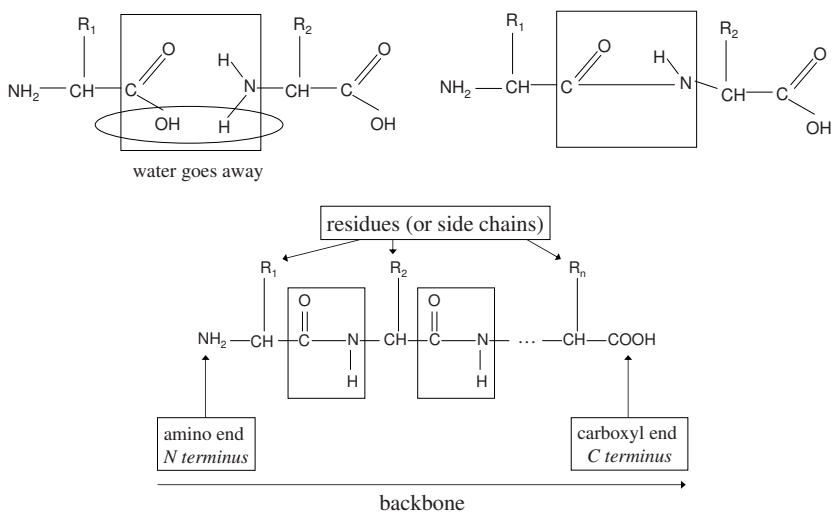
In this section I introduce a few things about the underlying biology, which are needed to get started. I then give the general form of the underlying model and discuss some of its attributes. In both cases, I elaborate in the next section in the context of each model.

2.1 Molecular Biology

All life depends upon three critical molecules:

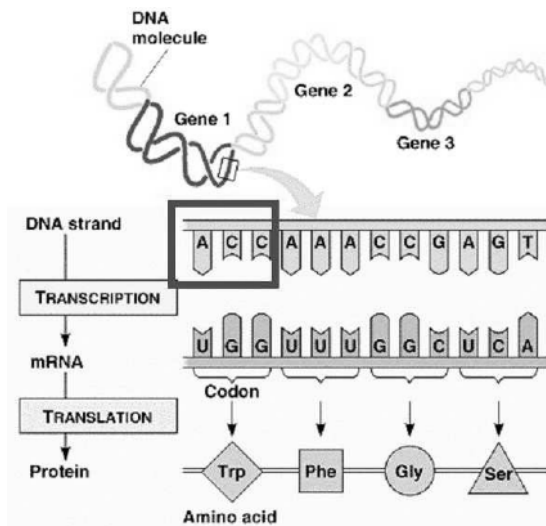
1. DNA, which contains information about how a cell works
2. RNA, which makes proteins and performs other functions
3. Proteins, which are regarded as the workers of the cell.

DNA is a double-stranded sequence of nucleic acids: Adenine, Cytosine, Guanine, and Thymine. RNA is a single-stranded sequence of nucleic acids: Adenine, Cytosine, Guanine, and Uracil. The genetic code maps each triple of nucleic acids into one of 20 amino acids. A *peptide bond* is a bonding of two amino acids. A protein is determined by a sequence of successively bonded amino acids. Figure 1 shows the

**Fig. 1.** Bonding Amino Acids

bonding of two amino acids — the carboxyl end of one bonds with the amino end of the other, discarding water. It then shows a generic backbone defined by a sequence of amino acids.

The *central dogma* is the information flow from DNA to proteins. (This use of “dogma” may seem strange — see [5, p. 116] for an explanation.) Figure 2 illustrates this. I shall say more as needed, in the context of each problem.

**Fig. 2.** The Central Dogma of Molecular Biology