

Sobre a modelagem da expressão gênica do vírus fago- λ através de equações diferenciais

Roseli Aparecida Brito Jordão

13 de julho de 2005

DISSERTAÇÃO APRESENTADA

AO

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

DA

UNIVERSIDADE DE SÃO PAULO

PARA

OBTENÇÃO DO GRAU

DE

MESTRE EM MATEMÁTICA APLICADA

Área de Concentração: Equações Diferenciais

Orientador: Prof. Dr. Eduardo Colli

Durante a realização deste trabalho a autora recebeu apoio financeiro do CNPq.

Agradecimentos

Em primeiro lugar, quero agradecer a DEUS por ser tão generoso comigo, pelas graças que tem me concedido e por não me abandonar nos momentos difíceis.

Aos meus pais, Gildásio e Gilza, pela minha existência e principalmente, à minha mãe, pelas horas de dedicação e cuidados ao meu mais precioso tesouro: meus filhos.

Aos meus filhos, Guilherme e Ana Giulia, minhas riquezas, por me proporcionarem momentos maravilhosos que me ajudaram a recuperar a energia. A eles e também aos meus alunos dedico este estudo, pois apesar de ainda pequenos, já observam que a dedicação aos estudos é um caminho que só traz benefícios e recursos para transformar a realidade.

Ao meu esposo, Dráuzio, pelo encorajamento e companheirismo.

A meus irmãos, Ailson, Gilberto, Rosiléa e Renata, pelo apoio e auxílio, com suas valiosas opiniões.

A minhas cunhadas, Dulcelene e Denise, pelas palavras de incentivo.

Ao professor Eduardo Colli, pela sua valiosa orientação, sem a qual não seria possível a realização deste trabalho.

Aos demais professores do IME-USP que contribuíram de alguma maneira na minha formação acadêmica.

Ao CNPq pelo apoio financeiro no início do programa de mestrado.

Aos muitos colegas e amigos, que fiz durante este curso e que me ajudaram comparti-

lhando conhecimentos. Não poderia deixar de citar de forma especial, Laura e Rita.

A todos que, de alguma maneira, ajudaram a concluir este trabalho. **Muito Obrigada.**

Sumário

1	Introdução	7
2	Descrição dos processos biológicos	11
2.1	O hospedeiro	11
2.2	O parasita	12
2.3	A infecção	13
2.4	O DNA	14
2.5	O DNA do fago λ	17
2.6	A regulação gênica	19
2.7	O controle	24
2.8	A indução por radiação ultravioleta	26
2.9	Descrição das proteínas regulatórias	29
2.10	Competição, interação e cooperatividade	32
3	Termodinâmica Estatística	35
3.1	Pressupostos provenientes da física quântica	35
3.2	Partícula numa caixa	38
3.3	Contagem dos níveis de energia	41
3.4	Pressupostos estatísticos da Termodinâmica	42

3.5	Interação entre sistemas macroscópicos	44
3.6	Formas diferenciais	46
3.7	Entropia e temperatura	48
3.8	Fator integrante	51
3.9	A função entropia e a relação de Euler	52
3.10	A energia livre de Gibbs	55
3.11	Equilíbrio químico	58
3.12	Dedução probabilística da equação de equilíbrio	61
4	O modelo	63
4.1	Probabilidades das configurações	63
4.2	Tabela de configurações	66
4.3	Equação diferencial	71
4.4	Comentários sobre o modelo original	75
4.5	Constantes utilizadas no modelo	78
5	Descrição da dinâmica do modelo	82
5.1	Análise do modelo	82
5.2	Região invariante	82
5.3	Encontrando singularidades	86

5.4	Classificação das singularidades	89
5.5	O retrato de fases	92
5.6	Comentários sobre os parâmetros	94
6	Considerações finais	100

1 Introdução

Este trabalho tem como base o modelo físico-químico formulado por Madeline A. Shea e Gary K. Ackers em seu artigo “*The control O_R system of Bacteriophage lambda - A physical-chemical model for gene regulation*”, que descreve uma parte do processo biológico da infecção sofrida pela bactéria *E.coli* quando em contato com um vírus denominado *fago* λ . O modelo é um sistema de equações diferenciais baseado nos processos de regulação gênica e na Termodinâmica Estatística.

Nosso objetivo é compreendê-lo através da descrição dos processos biológicos essenciais da regulação gênica, justificá-lo com os princípios da Termodinâmica Estatística e dar-lhe um enfoque matemático, explorando o sistema de equações diferenciais em \mathbb{R}^2 , simulando o modelo e discutindo os resultados.

Este estudo servirá para organizar de forma autocontida as informações necessárias para o entendimento desse tipo de modelagem. Servirá também para deixar mais claro que a existência de situações de equilíbrio pode ser naturalmente explicada pelas singularidades do sistema, mostrando também a importância de se considerar os dados bioquímicos sobre as taxas de degradação de proteínas, o que não foi feito no trabalho original.

Para introduzirmos o leitor no assunto, passamos a uma breve descrição do mesmo.

Existe naturalmente uma tendência da bactéria *Escherichia coli* (*E.coli*) ser infectada pelo bacteriófago (fago) λ - um vírus destruidor de bactérias. Esta característica foi des-

coberta no Instituto Pasteur, em Paris, há mais de 40 anos ([13]). Tal processo é bem interessante pois verificamos que uma bactéria pode ser infectada por um vírus e, a partir daí, ter seu metabolismo e suas funções modificados.

A bactéria *E.coli* vive pacificamente no intestino humano ([9]). Pelo seu metabolismo natural não causa nenhum mal ao seu hospedeiro (o homem), pelo contrário, pode até quebrar resíduos complexos e transformá-los em substâncias simples gerando vitaminas *D*, *K* e *B*₁₂. Além disso sua população é controlada pelo organismo humano. Porém, ao ser infectada por um determinado vírus torna-se tão agressiva que recebeu até um código de identificação: é a *E.coli* O157:H7, que causa diarreia hemorrágica, podendo levar seu hospedeiro (o homem) à morte.

Por outro lado, existe um exemplo benéfico. Esta mesma bactéria, ao ser infectada por um outro vírus, o T_7 recombinado com o λ , leva o intestino humano a produzir insulina, o que pode regularizar o organismo de uma pessoa diabética.

Podemos ver que o estudo deste tipo de infecção de uma bactéria por um vírus é relevante tanto para evitar o mal como para providenciar o bem. Em ambos os casos um modelo matemático qualitativo pode trazer contribuições valiosas, pois este pode prever o tipo de desenvolvimento de uma bactéria ao longo do tempo, através da avaliação da evolução das concentrações de proteínas específicas presentes em bactérias infectadas, dada uma concentração inicial das mesmas. Desta forma, poderemos conhecer as possibilidades

de desenvolvimento de uma colônia como um todo. O fago λ tem crescimento e controle semelhante aos dos vírus citados acima ([9]), então este tipo de modelagem caberá para ambos.

Falaremos primeiramente, no Capítulo 2, dos processos bioquímicos necessários para a modelagem e dos elementos biológicos que compõem o sistema, expondo que após a infecção duas situações podem ocorrer: instalar-se o estado lisogênico (“dormente”), ou o lítico (de reprodução do vírus e “destruição” da bactéria). Pretendemos utilizar o modelo do artigo original para mostrar que os dois estados podem ser vistos como dois equilíbrios do mesmo sistema de equações diferenciais.

Esta modelagem não inclui o momento de decisão inicial pois o modelo tem origem estatística e não se aplica bem aos instantes iniciais de infecção ([1]).

No Capítulo 3 faremos um breve resumo da Termodinâmica Estatística até o conceito de energia livre de Gibbs e exporemos sua relação com a equação de equilíbrio químico.

Esta equação de equilíbrio químico e sua relação com a energia livre de Gibbs foram utilizadas no Capítulo 4, juntamente com dados apresentados em [15], para a formulação do modelo quantitativo, através de um sistema de equações diferenciais em \mathbb{R}^2 .

No Capítulo 5 apresentamos a dinâmica do modelo e discutimos o seu retrato de fases, fazendo considerações qualitativas sobre o mesmo.

Encerramos este trabalho, no Capítulo 6, com várias considerações e pontos relevantes

para futuras discussões sobre este tipo de modelagem. Não faremos, no entanto, uma comparação numérica entre os dados oriundos do modelo apresentado e os dados reais vindos das experiências que comprovem a eficácia quantitativa da modelagem.

2 Descrição dos processos biológicos

2.1 O hospedeiro

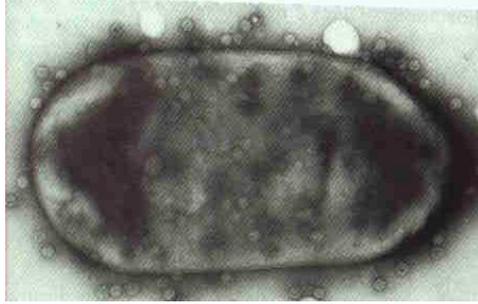


Figura 1: Foto de uma bactéria *E. coli* ([13])

A bactéria *Escherichia coli*, ou simplesmente *E.coli* - ver Figura 1 - é um indivíduo unicelular, procarionte, com núcleo simples contendo uma única molécula de DNA circular, e seu hialoplasma contém apenas ribossomos, que são organelas responsáveis pela produção de proteínas. Sua locomoção se dá através de flagelos e seus movimentos não são casuais nem caóticos. De acordo com [5] esta bactéria tem um tamanho aproximado de 2×10^{-6} m (metros) ou 2μ m de comprimento por aproximadamente 1μ m de diâmetro. Se imaginarmos a bactéria, aproximadamente como um cilindro temos que seu volume é de

$$V = \pi\left(\frac{1}{2}\right)^2 \times 2 = \frac{\pi}{2}\mu m^2 = \frac{\pi}{2} \times 10^{-15} \text{ litros.} \quad (1)$$

Este valor é utilizado em [15] e [10] com esta mesma finalidade.

Na Seção 4.5, veremos que este valor será útil para o cálculo da concentração, que é

dada em M (molar) = moles / litros.

A massa da *E. coli* é formada por: 70% de água, 15% de proteínas, 6% de moléculas de RNA, 3% de açúcares, 2% de gorduras, 1% da molécula de DNA, 1% de sais minerais e pequenas quantidades de outras substâncias. Seria necessária uma colônia de 1 trilhão de *E. coli* para alcançar 1 grama de massa ([9]).

2.2 O parasita

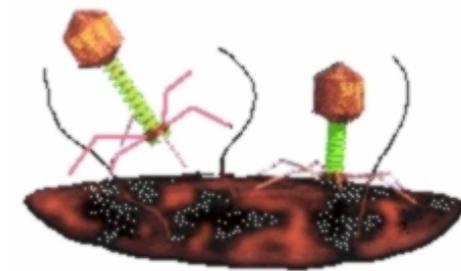


Figura 2: Ilustração do fago- λ atacando a bactéria

Bacteriófagos ou simplesmente **fagos** são vírus “comedores” de bactérias - ver Figura 2. O parasita fago λ , protagonista do nosso estudo, possui apenas uma molécula de DNA com fita dupla - existem fagos com fita simples de DNA e ainda com uma fita de RNA - e uma cápsula formada por 15 proteínas diferentes, com uma cabeça na forma de um icosaedro e uma cauda ([13]). É obrigatoriamente um parasita, pois para se multiplicar necessita utilizar os recursos de um outro ser vivo, neste caso os de uma bactéria.

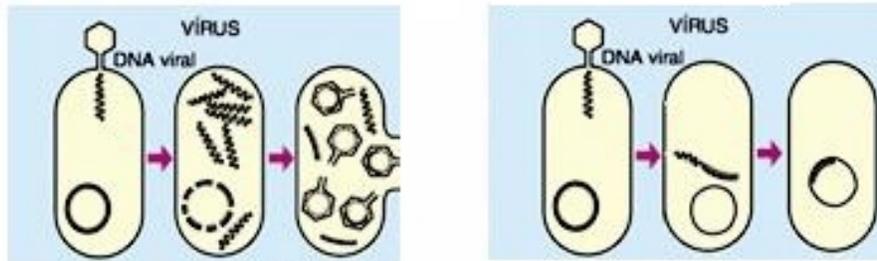


Figura 3: Momento da infecção ([13]). Estado lítico (à esquerda). Estado lisogênico (à direita).

2.3 A infecção

Quando o vírus entra em contato com a bactéria, ele se fixa na parede de sua célula e injeta para dentro da mesma seu DNA, descartando sua cápsula corporal de proteínas ([13]). Ao entrar na célula, o DNA do vírus se circulariza: uma enzima do hospedeiro liga covalentemente os terminais 3' e 5', cujos formatos são complementares (terminais coesivos).

O fago λ é dito **temperado** pois pode assumir dois estados distintos de crescimento após a infecção:

1. O estado **lítico**, também chamado de virulento, no qual o DNA do vírus (fago) toma o controle sobre o mecanismo da bactéria e começa a se reproduzir, replicando seu próprio código e fabricando novas cápsulas proteicas. Com isso a célula vai ficando repleta com vários vírus prontos até que a célula da bactéria *lise* (estoure). Esta informação pode nos

dar a impressão errada de que é a saturação da célula por novos vírus que a faz estourar, mas na verdade o estado lítico desencadeia uma série de processos que culminam na destruição da membrana celular. Este processo dura cerca de 45 minutos e o número de novos vírus por célula pode variar de 50 a 1000 indivíduos ([15]).

2. O estado **lisogênico**, no qual o DNA do fago integra-se ao da bactéria, permanecendo dormente. Neste caso passa a chamar-se *profago* e não faz mal algum à bactéria. Porém, quando ela se reproduz, gera nova bactérias já infectadas ([13]).

Ambos os processos de reprodução são eficientes pois tanto o vírus quanto a bactéria se reproduzem rapidamente gerando centenas de novos indivíduos em questão de minutos. De fato, é graças à rapidez com que os indivíduos envolvidos se reproduzem que se pode combinar genética com bioquímica para analisar a regulação gênica muito mais eficazmente do que em organismos superiores.

2.4 O DNA

O DNA é uma molécula inerte por si só, mas contém muitas informações a serem decodificadas. Essa molécula é composta por duas fitas formadas por uma seqüência de 4 bases hidrogenadas que se ligam 2 a 2 formando pontes que unem as duas fitas.

Cada uma destas bases tem uma estrutura particular, o que faz com que a molécula tenha uma superfície irregular, porém definida pela seqüência destas bases. Cada grupo de 3 bases codifica um dos 20 aminoácidos existentes. Todo aminoácido possui dois grupos

básicos em sua composição. Esses grupos também podem ser chamados de terminais, pois ficam nas extremidades dos aminoácidos, ou ainda, domínios. Um deles é chamado *Carboxyl*, com a composição química dada por $COOH$ - também chamado de *C-terminal*. O outro é denominado *Amino*, sua composição química é dada por NH_2 - também chamado de *N-terminal*. Existem algumas exceções, mas em geral, são estes terminais que permitem aos aminoácidos ligarem-se uns aos outros, de forma que o terminal amino de um liga-se ao terminal carboxyl do outro.

Os aminoácidos podem se juntar em pequenas ou grandes quantidades e ainda em diversas seqüências diferentes, formando assim as proteínas, cada uma com seu formato característico dependendo da seqüência e do número de aminoácidos que a compõe. Apresentam-se tipicamente como um glóbulo irregular cuja superfície é cheia de cavidades e protuberâncias. Esta característica está intimamente envolvida num processo denominado *regulação gênica*, que será explicado na Seção 2.6, pois há proteínas que se ligam em seqüências específicas do DNA ao longo de um subconjunto de pares de bases (pb), e cada uma seleciona seu(s) local(is) de ligação(ões) entre uns 5 milhões ou mais de pares de bases, pela complementaridade de suas estruturas. Esta característica então permitirá que o encaixe ocorra somente entre sítio (local do DNA) e proteína específicos.

A estrutura da proteína liga-se ao DNA somente sobre um determinado composto e efetua sempre o mesmo tipo de reação. Esta grande especificidade está relacionada à forma

tridimensional de ambos. Eles se encaixam perfeitamente como chaves em fechaduras, por isso este processo é chamado de *Teoria da Chave-Fechadura*.

No processo de *replicação*, a molécula de DNA atrai para si uma outra chamada RNA polimerase (RNAP), que é uma enzima - uma proteína especial. A função da RNAP é traduzir o código genético do DNA e transcrevê-lo numa nova molécula, que vai sendo formada no decorrer deste processo, o RNA mensageiro (RNAm). Ele funcionará como um negativo de uma das fitas do DNA e levará a informação para o ribossomo, que é responsável por executar a transformação das trincas de pb em aminoácidos, produzindo assim as proteínas.

A RNAP liga-se num determinado ponto próximo do início de um gene, num local chamado de *sítio promotor*, e viaja ao longo de um ou mais genes sintetizando o RNAm. Ocorre continuamente um desenrolamento e um reenrolamento de pequenos trechos do DNA, que se separam temporariamente para a passagem da RNAP e a formação do RNAm. Esta mudança temporária na característica do DNA, é chamada de *formação de complexo aberto*.

Cada fita do DNA tem um sentido específico de transcrição, como se fosse uma rodovia duplicada, onde cada pista pode ser trafegada num único sentido, isto é, cada fita só pode ser copiada num sentido. Cada promotor aponta a RNAP num ou noutro sentido ao longo da hélice do DNA. Assim copia apenas uma das fitas do DNA de cada vez.

2.5 O DNA do fago λ

De forma geral, o λ -DNA tem dois genes, que podem ser considerados, mais importantes. São os genes *cro* e *cI*. Estes genes são adjacentes no λ -DNA, *mas seus promotores apontam a RNAP em sentidos opostos*.

De acordo com [13], no estado lisogênico, praticamente só o gene *cI* do profago é transcrito. Esta característica é incorporada no modelo, prevendo que o gene *cro* tenha pouca chance de ser transcrito no estado lisogênico.

A transcrição deste gene pela enzima RNAP (RNA polimerase) produzirá um RNAm (RNA mensageiro) que ao ser traduzido sintetizará a proteína chamada *CI-Repressor* ou *Rep*, responsável pela manutenção do estado lisogênico. Esta proteína reprime a transcrição do outro gene, o *cro*, o qual só poderá ser transcrito eficazmente na indução para o estado lítico.

A transcrição deste outro gene gera a proteína Cro, que reprimirá a síntese da proteína Rep e permitirá o avanço do crescimento lítico.

A Figura 4 mostra um esquema do mapa genético do λ -DNA. Seu genoma completo envolve uns 50000 pb. O vírus carrega genes que são responsáveis por várias funções: gerar enzimas que permitirão encaixar seu DNA no DNA da bactéria, gerar enzimas que farão o contrário, se ocorrer uma mudança para o estado lítico, gerar a cápsula de proteínas para a formação de sua estrutura, etc. A produção de proteínas a partir desses genes é feita com

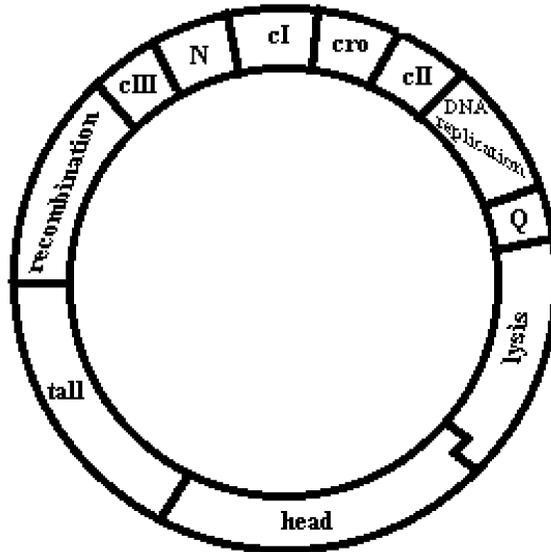


Figura 4: Mapa genético do fago λ .

os recursos da célula invadida.

Os outros genes pertencentes ao λ -DNA, não participam da modelagem mas têm seus papéis definidos em cada estado de crescimento da bactéria. Os genes *N*, *cII* e *cIII* atuam desde o momento da infecção e contribuem na decisão inicial sobre qual caminho seguir. Se a decisão for pelo crescimento lítico, passam a atuar os genes *Q*, *cro*, *DNA replication*, além de *N*. Em seguida, mais genes também são *ativados*, os dos grupos *head* e *tail* e por fim os do grupo *lysis*.

No outro caminho, para o crescimento lisogênico, os genes que atuam são: *cI*, *cII* e *recombination*.

A Tabela 1 apresenta de forma esquemática os genes do λ -DNA e suas funções.

GENE	AÇÃO
	Genes importantes:
<i>cI</i>	gene que sintetiza a proteína Rep.
<i>cII</i>	gene que sintetiza a proteína CII.
<i>cIII</i>	gene que sintetiza a proteína CIII.
<i>cro</i>	gene que sintetiza a proteína Cro.
<i>N</i>	gene que sintetiza a proteína N.
<i>Q</i>	gene que sintetiza a proteína Q .
	Grupos de genes reunidos pela função:
<i>recombination</i>	10 genes que sintetizam proteínas que ligam os DNA's.
<i>DNA replication</i>	2 genes que sintetizam proteínas que duplicam o DNA.
<i>lysis</i>	3 genes que sintetizam proteínas que controlam o crescimento lítico.
<i>head</i>	~ 10 genes que fabricam as cabeças das cápsulas de proteínas.
<i>tall</i>	~ 12 genes que fabricam as caudas das cápsulas.

Tabela 1: Descrição do mapa genético

2.6 A regulação gênica

A regulação gênica é um processo biológico muito simples, porém, fundamental neste e em outros casos e ocorre naturalmente desde o início da vida. Por exemplo, quando um ser vivo está em formação, tudo começa a partir de um ovo fertilizado, gerando células que vão se dividindo milhões de vezes, onde cada nova célula recebe um conjunto de genes idêntico. No decorrer do processo de desenvolvimento vários tipos de células vão formando as diversas partes do corpo deste novo ser, porém todas elas têm o mesmo código genético.

O que vai determinar o formato, a função e outras características de cada célula é a

relação dos genes *ativos* e *inativos* em seu DNA e ainda o grau de atividade dentro da célula envolvendo a transcrição destes genes, ou seja, cada célula usa só um subconjunto de seus genes para a produção direta de outras moléculas que podem inclusive influenciar nas transcrições de genes. Dizemos que a relação destes genes e os graus de atividade envolvendo suas transcrições é a *expressão gênica*.

Entendemos por genes *ativos*, aqueles que podem ser transcritos e *inativos* aqueles bloqueados e que portanto não podem ser transcritos. Cada gene pode estar temporariamente *ativo* ou *inativo*. E ainda, os genes *ativos* podem ter diferentes graus de atividade.

Um gene ativo, ao ser transcrito permite a produção de uma ou mais moléculas de proteínas. Estas proteínas poderão influenciar na transcrição de um ou mais genes, inclusive do próprio gene que as originou. Essa influência poderá gerar diferentes graus de atividades para os diversos genes ativos e ainda ativar ou desativar genes. As etapas da regulação gênica vão se desenrolando como uma cascata, um evento após o outro, de maneira que uma ordem seja estabelecida e que cada evento ocorra somente no momento adequado. Alguns genes são transcritos somente depois de algum tempo ou de algum comando. Por exemplo se o gene que produz o *lise* na bactéria fosse transcrito desde o início do processo, ela estouraria antes que novos vírus fossem produzidos.

Este processo não ocorre somente na formação do ser vivo, mas durante toda a vida da célula. Genes ativos podem tornar-se inativos e vice-versa, basta que algum fator interno ou

externo contribua para isso. Por exemplo, quando uma pessoa toma sol sua pele escurece porque os genes que aumentam a pigmentação nas células são ativados, ou seja, existem fatores que modificam o estado do gene. Por isso dizemos que a expressão gênica é *regulada*.

Os biólogos ainda estudam como ocorrem estas ativações naturais de genes durante o desenvolvimento do ser e ao longo de sua vida. Procuram conhecer os mecanismos de regulação gênica que participam do circuito de decisão, quais garantem a ativação ordenada de um conjunto de genes, quais destes passos são dados por programas intracelulares e ainda quais são dados por sinais extracelulares.

A regulação gênica é um processo que se baseia no seguinte fato: se o sítio promotor de um gene ou grupo de genes está livre, o gene está ativo e poderá ser transcrito assim que a RNAP se ligar ao promotor. E estará inativo se parte deste promotor estiver bloqueado por uma proteína. Este tipo de proteína é chamada *regulatória* justamente porque regula a ativação de um ou mais genes. Fica claro que estas proteínas exercem um *controle* sobre o processo: dizemos que este controle é *negativo* quando uma proteína se liga ao DNA, bloqueando um promotor e portanto impedindo que um ou mais genes sejam transcritos, e dizemos que é *positivo* quando uma proteína se liga ao DNA num local adjacente ao promotor de forma a atrair mais fortemente a RNAP para aquele promotor e a aumentar a chance de que um ou mais genes sejam transcritos.

De um modo geral, o local do DNA onde as proteínas regulatórias se ligam chama-se

sítio operador e estende-se por aproximadamente 17 pares de bases, podendo tomar parte de 1 ou 2 promotores.

Um tipo de proteína regulatória em particular se encaixa num ou mais sítios operadores na molécula de DNA. Um sítio operador poderá ser **direito** ou **esquerdo** indicando o sentido para o qual ocorre a transcrição dos genes da fita do DNA em que está. Porém, mesmo sendo um operador direito, por exemplo, ao ser ocupado poderá bloquear um gene cujo promotor é esquerdo, isso porque as fitas do DNA formam uma espécie de hélice e as proteínas, com sua forma tridimensional podem, ao se encaixarem num operador direito, encobrir parcialmente um promotor esquerdo.

Portanto, se a proteína estiver presente, usualmente ela será encontrada no sítio operador, mas ela pode facilmente sair do sítio, permitindo assim que outra proteína idêntica se encaixe. Tudo irá depender da sua *concentração* - dada em M (molar) = moles/litro - e da sua *afinidade* - força com que o encaixe ocorre - pelo sítio operador. As ligações nunca são definitivas: ligamentos e desligamentos ocorrem a todo instante, e é nesse fato que se baseia a modelagem que apresentaremos.

Os sítios operadores nos quais as proteínas regulatórias Rep e Cro podem se encaixar no λ -DNA são três, todos operadores direitos, denominados O_{R1} , O_{R2} e O_{R3} . Eles ficam em locais adjacentes, não se sobrepõem e tomam parte dos promotores dos genes que ativam a síntese destas proteínas.

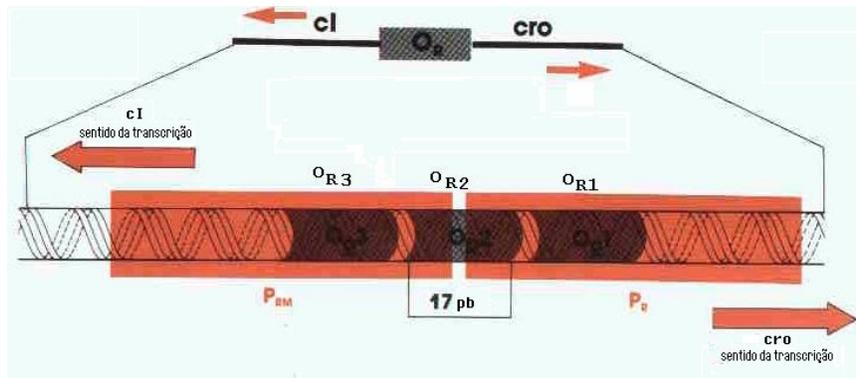


Figura 5: Ilustração dos sítios promotores e operadores ([13]).

Como podemos ver na Figura 5 os sítios operadores O_{R1} e O_{R2} tomam parte do promotor do gene *cro* denominado P_R e os sítios operadores O_{R2} e O_{R3} tomam parte do promotor do gene *cI* denominado P_{RM} .

Para cada sítio operador existe uma força de encaixe, denominada *afinidade* com a proteína e esta pode variar dependendo da situação. Se um sítio tem uma afinidade menor que a do outro então em algum instante de baixa concentração da proteína somente o sítio com maior afinidade estará ocupado. Mas uma alta concentração da proteína poderá fazer com que a diferença de afinidade seja ignorada e assim ambos serão ocupados. **As ligações entre proteínas e sítios operadores são muito importantes, pois determinam se um ou mais genes estão ou não ativos** ([13]).

Nas seções 2.7 e 2.10 descreveremos qualitativamente duas situações:

✓ A manutenção do estado lisogênico → estando a bactéria lisogênica, ou seja, com

o profago integrado em seu DNA, determinaremos quais os níveis de concentrações das proteínas que garantem a manutenção da atual situação.

✓ A mudança para o estado lítico → situação contrária à anterior, ou seja, quais alterações nos níveis destas concentrações acarretam a mudança de estado.

Em resumo, sabemos que dois genes presentes no DNA do fago são os mais importantes e de forma bem simplificada são os responsáveis por ambas as situações (Tabela 2).

GENE	PROTEÍNA	AÇÃO
<i>cI</i>	CI-Repressor ou Rep	Responsável pela manutenção do estado lisogênico.
<i>cro</i>	Cro	Responsável pelo crescimento lítico.

Tabela 2: Ação das proteínas regulatórias

Um esclarecimento sobre a notação: Como em todas as nossa referências, utilizaremos letra inicial minúscula e caracteres em itálico para representar os genes *cro* e *cI* e letra inicial maiúscula para representar as proteínas Cro e Rep.

2.7 O controle

A decisão inicial sobre qual caminho de reprodução seguir, que ocorre logo após a infecção, se dará exclusivamente no DNA do fago, também através da *regulação gênica*. Há aqui uma questão que parece ainda não tão bem compreendida, sobre a decisão inicial. Em [13] e [15] trata-se de um processo de regulação gênica, assim como todas as decisões e processos posteriores, porém envolve mais elementos (genes, proteínas, etc). Já em [1], esta é tratada

com a adição de flutuações probabilísticas, que podem levar a um ou outro estado.

Esta decisão também depende de fatores externos, tais como a condição nutricional da célula e do meio em que ela está. Se decidir pelo crescimento lítico, o caminho é praticamente irreversível ([13]) até o estouro da bactéria, mas se optar pelo crescimento lisogênico, temos interesse em estudar quais fatores são cruciais para a manutenção deste estado e quais concentrações das proteínas provocariam uma mudança posterior para o estado lítico. Teoricamente, verificaremos no modelo que há possibilidade de reverter o estado lítico introduzindo-se uma quantidade de proteínas Rep. Mas numa situação real, ao entrar no estado lítico, muita coisa muda no interior da bactéria: o DNA do fago se separa do DNA da bactéria, mais vírus e portanto mais material genético é fabricado e a célula acaba por estourar, de forma que nosso modelo teórico perde a validade.

No estado lisogênico, *cI* é o gene do profago com maior grau de transcrição e, portanto, aquele que sintetiza mais proteínas. Este gene pode ser regulado tanto positiva quanto negativamente na expressão gênica. Estando ativo, sintetiza a proteína Rep, que se liga ao DNA do profago e ajuda a sintetizar mais proteínas do mesmo tipo (controle positivo). Portanto, contribui no aumento de sua própria concentração dentro da célula. Essa proteína também se liga ao λ -DNA, bloqueando o sítio promotor do gene *cro*, tornando-o inativo (controle negativo). Além disso, as proteínas livres também podem se ligar a outros λ -DNA's que venham a reinfectar a bactéria, tornando-os inativos (controle negativo). Isto

mostra que outro fago não pode começar o ciclo lítico numa bactéria que já esteja no estado lisogênico. Diz-se que a *célula lisogênica* é imune à infecção por um λ -vírus ([13]).

2.8 A indução por radiação ultravioleta

No estado lisogênico, Rep entra em O_{R1} e O_{R2} (esta situação juntamente com a RNAP no sítio promotor P_{RM} é o estado *transcricional* mais provável do sistema no geral) bloqueando P_R (desativando *cro*) e estimulando a transcrição de *cI*. Nessa situação, a proteína Rep está sendo sintetizada em alta escala. Assim a célula se reproduz (cresce e se divide), com o gene *cro* permanecendo com um grau de transcrição baixíssimo. Aqui, assim como em [15], não nos preocuparemos com as variações de volume e a presença de 2 DNA's nos momentos que antecedem a divisão celular, mas tal informação poderá ser incorporada em trabalhos complementares se houver a pretensão de se simular um modelo mais realista. Se a concentração de Rep aumenta muito, então o sítio operador O_{R3} será também ocupado, diminuindo drasticamente a produção de Rep (auto-regulação - que poderá ser confirmada no modelo). Porém na divisão celular a concentração cai e O_{R3} é novamente desocupado, portanto *cI* volta a ser ativado. Embora nosso modelo não preveja a divisão celular, outro mecanismo de diminuição da quantidade de proteína Rep é a degradação natural que ela sofre, assim como qualquer outra proteína. É isso que permite atingir-se um equilíbrio na concentração de Rep.

Numa experiência ([13] e referências ali contidas) realizada na década de 50, uma po-

pulação de bactérias infectadas foi exposta a uma certa quantidade de radiação ultra-violeta e os *profagos* tornaram-se virulentos, ou seja, passaram do estado lisogênico para o estado lítico. Muitos agentes semelhantes ao raio UV induzem a mudança do ciclo lisogênico para o lítico, matando a bactéria. Em [13] é cogitada a hipótese de que o raio UV danifique o DNA da bactéria o que irá levá-la a morte, então o profago sente através de enzimas bacterianas o iminente falecimento de seu anfitrião e muda para o ciclo lítico. Mas existe ainda uma outra hipótese dada em [13] e [15] de que, com a irradiação, a proteína regulatória que é responsável pela manutenção do estado lisogênico começa a ser degradada e então permite o avanço do crescimento lítico. É bom salientar que, mesmo na ausência de fatores externos, existem casos isolados de bactérias que entram espontaneamente no estado lítico.

Ficaremos com a hipótese de que, com a radiação ultravioleta, a proteína Rep é destruída (aumenta o grau de degradação). Essa maior degradação ocorrerá porque uma proteína bacteriana chamada RecA, que em estado normal catalisa recombinações entre moléculas de DNA, fica hiperativada e passa a quebrar moléculas de Rep ligadas ou não ao DNA.

A queda na concentração da proteína Rep se acentua e a produção da mesma também diminui, pois *cI* está quase inativo. Os sítios O_{R1} e O_{R2} permanecem mais tempo desocupados, o que diminui ainda mais a chance de transcrição de *cI* e a síntese da proteína Rep. Aumenta ainda mais a chance de RNAP ocupar o promotor P_R e aumentar o grau de transcrição do gene *cro*. O resultado disso é que a proteína Cro passa a ser sintetizada mais

eficazmente e tem mais chance de se ligar ao DNA do fago, porém com efeitos contrários. Com o aumento da sua concentração, a proteína Cro é que determina o curso agora. Neste ponto o “interruptor” é mudado e começa o estado lítico.

A ação da Cro é mais simples que a da Rep pois esta se liga independentemente nos O_R 's (não se ajudam) e portanto o controle sobre as transcrições é estritamente negativo - quando bloqueia P_{RM} , não permitindo a transcrição de cI . Suas afinidades com os sítios operadores são opostas às da Rep, como mostra a Tabela 4.

Conforme aumenta a concentração da Cro, aumentam as chances destas se ligarem nos sítios operadores. Se a concentração ficar muito alta, esta proteína pode bloquear seu próprio promotor e impedir sua própria transcrição (auto-regulação). Mas o estado lítico e as ações necessárias para a reprodução de novos vírus já estão em pleno andamento e como já dissemos é um processo praticamente irreversível.

Resumidamente, a constituição do controle (interruptor) gênico é dada pelas duas proteínas regulatórias, Rep e Cro que competem na ocupação de três sítios operadores direitos, O_{R1} , O_{R2} , O_{R3} e a enzima RNAP que poderá se ligar em dois sítios promotores P_R e P_{RM} .

Podemos dizer, aproximadamente, que o interruptor tem duas posições: no ciclo lisogênico temos cI **ativo** e cro **inativo**, enquanto que no ciclo lítico cI **está inativo** enquanto cro **está ativo** (Tabela 3).

CICLO LISOGÊNICO	<i>cI</i> on e <i>cro</i> off
CICLO LITICO	<i>cI</i> off e <i>cro</i> on

Tabela 3: Ativação dos genes

Estes 3 sítios operadores são semelhantes porém não idênticos, assim podem ser diferenciados pelas proteínas regulatórias. Isso explica as diferentes afinidades que as proteínas regulatórias têm com estes locais (Tabela 4).

Rep	O_{R1}	O_{R2}	O_{R3}
afinidades	forte	média	fraca
Cro	O_{R1}	O_{R2}	O_{R3}
afinidades	fraca	média	forte

Tabela 4: Intensidade das afinidades

A RNAP é provida pelo anfitrião e, quando se liga em P_R , transcreve para a direita o gene *cro*, e ao ligar-se em P_{RM} transcreve para a esquerda o gene *cI*. Não ocorrerá portanto, transcrição simultânea por uma única RNAP de ambos os genes, pois as mesmas ocorrem em sentidos contrários.

2.9 Descrição das proteínas regulatórias

A forma básica na qual a proteína Rep é sintetizada se constitui de dois domínios ou terminais ligados por uma espécie de haste, como mostra a Figura 6. Um destes terminais é denominado Carboxyl ($COOH$) pela composição química da extremidade livre do seu último aminoácido. Pelo mesmo motivo o outro domínio é chamado Amino (NH_2). A

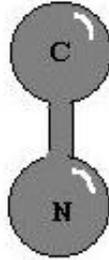


Figura 6: Molécula de monômeros da proteína Rep ([13]). *C* domínio Carboxyl. *N* domínio Amino.

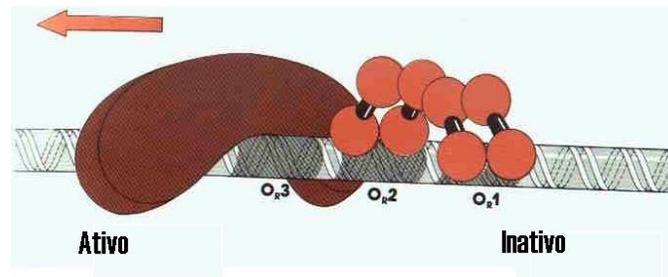


Figura 7: Encaixe dos dímeros de Rep nos sítios operadores O_{R2} e O_{R3} tornando o gene *cro* inativo e RNAP no sítio promotor P_{RM} tornando o gene *cI* mais ativo ([13]).

estrutura toda é composta por 236 aminoácidos, onde 92 compõem o domínio Amino, 104 compõem o domínio Carboxyl e 40 formam uma espécie de haste que une os dois domínios.

Esta forma simples na qual a proteína é sintetizada denomina-se monômero, porém para ligar-se ao DNA e cumprir o seu papel de proteína regulatória deve juntar-se a outra idêntica e formar um dímero (dupla de monômeros). O contato entre os monômeros ao formar dímeros é feito pelo domínio Carboxyl enquanto que o encaixe no DNA se dá através do domínio Amino, ver Figura 7.

Em nossas referências, existe uma divergência nas proporções entre número de dímeros e número de monômeros da proteína Rep presente na célula lisogênica.

Em [13] temos que praticamente 95% das Rep estão formando dímeros. Já em [15] esta proporção é dada como de 80%. Nosso modelo utiliza a primeira hipótese.

Desta forma, se é dada uma população inicial de 200 moléculas de monômeros de Rep, temos 95 moléculas de dímeros e 10 moléculas de monômeros livres na célula. A relação entre o número de moléculas de dímeros de Rep é de 47.5% do total de moléculas de monômeros presentes.

A proteína Cro também é sintetizada na forma simples de monômeros. Estes se ligam formando dímeros para o encaixe no DNA, como vemos na Figura 8, começando pelo sítio O_{R3} . Quando a concentração de dímeros de Cro é muito alta, esta proteína ocupa o sítio operador O_{R1} e pode bloquear sua própria transcrição.

Estes monômeros possuem um único domínio responsável pelas duas interações e praticamente 100% formam dímeros ([13] e [15]).

Podemos dizer que existe um tipo de interação entre os monômeros destas duas proteínas, fazendo com que seja mais comum encontrá-las em forma de dímeros do que em forma de monômeros.

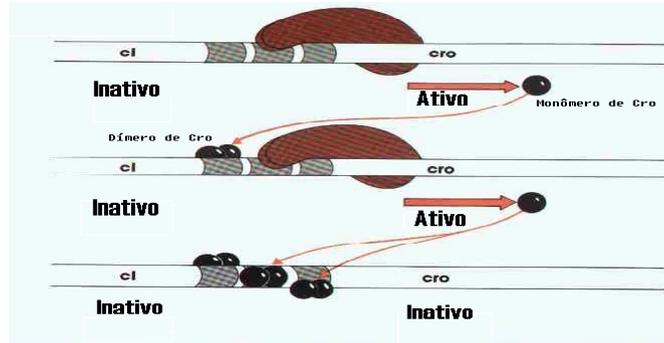


Figura 8: A RNAP em P_R , transcrição do gene *cro* ativa ([13]).

2.10 Competição, interação e cooperatividade

Tanto Cro quanto Rep formam dímeros para ligar-se em O_{R1} , O_{R2} e O_{R3} e ambas se encaixam nos mesmos locais no DNA, o que exclui a possibilidade de ambas ocuparem o mesmo sítio operador. Este fato nos dá uma idéia de competição entre elas. Estas ligações de natureza química não são covalentes e portanto podem ocorrer dissociações.

As possibilidades de ocupação nos sítios operadores pelas proteínas regulatórias e dos sítios promotores pela RNAP, assim como as características provenientes de cada uma destas possibilidades geraram uma série de afirmações que serviram de roteiro para a formulação do modelo físico-químico do nosso texto base ([15]).

Além disso, há vários tipos de interações importantes que devem ser e foram levadas em consideração, pois a força de atração entre proteína e DNA muda de intensidade devido à existência de algumas cooperações.

O dímero de Rep tem afinidades diferentes para cada O_R porque estes sítios operadores são semelhantes mas não idênticos. Estas pequenas diferenças fazem com que o dímero de Rep reconheça como preferencial o sítio O_{R1} . Normalmente após tê-lo ocupado o próximo será O_{R2} e, por último, se a concentração for suficientemente alta, ocupará O_{R3} .

Os dímeros de Rep também têm afinidades aos pares, ou seja, dois dímeros de Rep ligados em sítios operadores vizinhos, O_{R1} e O_{R2} ou O_{R2} e O_{R3} se ajudam mutuamente.

Essas afinidades são expressas em termos de energia livre, como veremos na Seção 3.11. Algumas das configurações dadas em [15] que podemos ver a título de exemplo são:

- ✓ Apenas um dímero da proteína Cro em O_{R3} .
- ✓ Apenas um dímero da proteína Cro em O_{R2} .
- ✓ Dois dímeros da proteína Cro ocupando simultaneamente O_{R3} e O_{R2} .
- ✓ Apenas um dímero da proteína Rep em O_{R1} .
- ✓ Apenas um dímero da proteína Rep em O_{R2} .
- ✓ Dois dímeros da proteína Rep ocupando simultaneamente O_{R1} e O_{R2} .
- ✓ Um dímero da proteína Rep em O_{R1} e um dímero de Cro em O_{R2} .

Todas as possibilidades com a quantificação das afinidades são dadas na Tabela 6, que será explicada oportunamente (Seção 4.2).

A posição mais importante para Rep é a ocupação de O_{R2} onde exerce controles positivo e negativo especificados na Tabela 5. Estando nesta posição o Rep aumenta a possibilidade

Positivo	quando ajuda a RNAP em P_{RM} , ativando a transcrição do gene cI
Negativo	quando impede a RNAP de entrar em P_R , desativando o gene cro

Tabela 5: Tipos de controle

de transcrição de cI , pois aumenta a chance da RNAP ligar-se em P_{RM} . Isso porque o domínio amino toca a RNAP fazendo mais uma interação - além da que já existe entre DNA e proteína - agora entre proteínas.

As afirmações mencionadas nos permitem escrever 40 configurações possíveis para as ligações envolvendo as proteínas regulatórias e os sítios operadores O_R 's e ainda entre a RNAP e os sítios promotores. Cada uma delas possível e mais ou menos provável, de acordo com suas afinidades e a concentração das proteínas regulatórias e ainda da RNAP. Mais adiante (Capítulo 4) veremos como usar estas informações para a construção do modelo matemático que prevê a dependência das concentrações em relação ao tempo.

Nesta etapa, já consideramos as características biológicas do nosso problema, porém na formulação do modelo são consideradas muitas características físicas e químicas que podem ser melhor entendidas através da Termodinâmica Estatística. Por isso devotaremos o próximo Capítulo ao entendimento de aspectos básicos desta teoria para depois voltarmos à quantificação dos processos descritos neste Capítulo.

3 Termodinâmica Estatística

A modelagem propriamente dita dos processos biológicos descritos no Capítulo 2 será feita apenas no Capítulo 4. Aqui apresentaremos de forma sucinta a teoria da Termodinâmica Estatística de forma a justificar a equação de equilíbrio químico (10) e a relação da constante de equilíbrio com a energia livre de Gibbs.

Neste Capítulo buscamos ainda entender o conceito de energia livre de Gibbs e, ao olharmos para a natureza química destas ligações, compreender como esta energia influencia na frequência de ocupação de cada sítio operador no λ -DNA pelas proteínas regulatórias e dos sítios promotores pela RNAP. Nosso intuito é dar ao leitor não iniciado as idéias básicas que estão por trás dos conceitos utilizados na formulação do modelo, para que a leitura seja praticamente autocontida.

Se ao leitor for familiar à teoria ou simplesmente assim o desejar, pode pular este Capítulo, admitindo-se a validade da Equação de equilíbrio (10), a qual retomamos e utilizamos no Capítulo 4.

3.1 Pressupostos provenientes da física quântica

Na Mecânica Quântica ([4]), a evolução de uma partícula é descrita por uma função $\psi(x, t)$ com $x \in \mathbb{R}^3$, $t \in \mathbb{R}$, a valores complexos. Essa função satisfaz a equação de Schrödinger

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \psi + V(x)\psi . \quad (2)$$

onde

$$\nabla^2\psi = \frac{\partial^2\psi}{\partial x_1^2} + \frac{\partial^2\psi}{\partial x_2^2} + \frac{\partial^2\psi}{\partial x_3^2},$$

onde $x = (x_1, x_2, x_3)$, m é a massa da partícula e $\hbar = \frac{6.626 \times 10^{-34}}{2\pi}$ (é a constante de Planck h dividida por 2π). Se adequadamente normalizada para que sua integral espacial valha 1, a função $|\psi|^2$ fornece a distribuição de probabilidade de se encontrar a partícula em determinada posição no instante t . Além disso, a função $V(x)$ é a energia potencial clássica associada a um campo de forças conservativo.

Se buscarmos soluções do tipo $\psi(x, t) = T(t)u(x)$ (com variáveis separáveis) com $T : \mathbb{R} \rightarrow \mathbb{C}$ e $u : \mathbb{R}^3 \rightarrow \mathbb{C}$, obteremos:

$$i\hbar \frac{dT}{dt}(t)u(x) = -\frac{\hbar^2}{2m} \nabla^2 u(x)T(t) + V(x)T(t)u(x)$$

e dividindo a equação toda por $T(t)u(x)$ teremos:

$$i\hbar \frac{T'(t)}{T(t)} = -\frac{\hbar^2 \nabla^2 u(x)}{2mu(x)} + V(x).$$

Para que haja a igualdade para todo (x, t) é preciso que exista uma constante $E \in \mathbb{C}$ tal que

$$i\hbar \frac{T'(t)}{T(t)} = E. \tag{3}$$

e

$$-\frac{\hbar^2 \nabla^2 u(x)}{2mu(x)} + V(x) = E. \tag{4}$$

A Equação (3) é ordinária e linear, e tem soluções do tipo $T(t) = Ce^{-\frac{iE}{\hbar}t}$. No caso da Equação (4), podemos reescrevê-la como:

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V(x) \right)u = Eu. \quad (5)$$

O operador $H : u \rightarrow \left(-\frac{\hbar^2}{2m}\nabla^2 + V(x) \right)u$ é chamado de *operador Hamiltoniano*. As soluções de (5) dependem de $V(x)$. Para entendê-la, observamos que H é um operador linear agindo num espaço de funções, em geral o espaço de Hilbert das funções L^2 (a derivação é entendida no sentido de distribuições). Se existem E e u tais que $Hu = Eu$ então u é autofunção para H e E é um autovalor.

O estudo de H deve ser feito caso a caso, mas admitiremos que tenha as seguintes propriedades, em geral verificadas nos casos de interesse:

(a) seu espectro é discreto, real, positivo e enumerável (isto é, é o conjunto de seus autovalores);

(b) o conjunto das autofunções é uma base ortonormal das funções L^2 .

Os autovalores E_1, E_2, E_3, \dots são chamados de *valores de energia* e o espectro é chamado de *espectro de energia*.

Como conseqüência, suponha que procuremos uma solução da equação de Schrödinger tal que $\psi(x, 0) = u_0(x)$. Então escrevemos $\psi(x, 0) = \sum_j a_j u_j(x)$ onde u_1, u_2, \dots são as

autofunções associadas aos autovetores. Segue que a função

$$\psi(x, t) = \sum_j a_j e^{-i\frac{E_j}{\hbar}t} u_j(x)$$

é a solução da equação de Schrödinger com a condição inicial prescrita, usando o fato de que este operador é linear.

Como $|\psi|^2$ fornece a distribuição de probabilidade em cada instante da posição da partícula sua integral deve valer 1 (na verdade após normalização apropriada). Por outro lado

$$\|\psi\|^2 = \int |\psi|^2 dx = \sum_j |a_j|^2,$$

assim a soma do lado direito deve ser igual a 1. A interpretação física é que $|a_j|^2$ representa a probabilidade de se encontrar a partícula no estado de energia E_j .

3.2 Partícula numa caixa

Um dos exemplos mais simples, mas que nos interessa bastante, é o de uma partícula numa caixa unidimensional, cujo potencial pode ser modelado por

$$V(x) = \begin{cases} 0 & \text{se } |x| < a \\ \infty & \text{se } |x| \geq a \end{cases}$$

Isso impõe que para x fora de $(-a, a)$ vale $\left(-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x)\right)u = Eu$ com $V(x) = \infty$, então $u(x) = 0$. Para x dentro de $(-a, a)$ vale

$$u''(x) + \frac{2m}{\hbar^2} Eu(x) = 0.$$

Se $E < 0$, temos a equação $u''(x) = k^2 u(x)$ cujo polinômio característico tem como raízes os valores k e $-k$. A solução geral desta equação é $u(x) = c_1 e^{kx} + c_2 e^{-kx}$ mas para esta solução se anular em $-a$ e a será preciso que $c_1 = c_2 = 0$ então $u(x) = 0$.

Se $E = 0$, temos $\frac{d^2 u}{dx^2}(x) = 0$, cuja solução geral é $u(x) = c_1 x + c_2$. Mas para esta solução se anular em $-a$ e a temos que $c_1 = c_2 = 0$ então, $u(x) = 0$.

Se $E > 0$, temos a equação $u''(x) = -k^2 u(x)$ cujo polinômio característico tem como raízes os valores ik e $-ik$. A solução geral desta equação é $u(x) = c_1 \cos(kx) + c_2 \sin(kx)$ com $k^2 = \frac{2mE}{\hbar^2}$.

As soluções que buscamos são combinações lineares de $\sin(kx)$ e $\cos(kx)$.

No caso dos senos buscamos $\sin(ka) = 0 = \sin(-ka)$. Isso implica que $ka = n\pi$ com $n = 1, 2, 3, \dots$, isto é,

$$k^2 = \frac{2mE}{\hbar^2} = \frac{n^2 \pi^2}{a^2}$$

ou ainda

$$E = \frac{(\hbar\pi)^2}{2ma^2} n^2$$

No caso dos cossenos $\cos(ka) = 0 = \cos(-ka)$. Isso implica que $ka = (n + \frac{1}{2})\pi$, isto é,

$\frac{2mE}{\hbar^2} = k^2 = (n + \frac{1}{2})^2 \frac{\pi^2}{a^2}$, então

$$E = \frac{\hbar^2 \pi^2}{2ma^2} (n + \frac{1}{2})^2$$

com $n = 0, 1, 2, 3, \dots$

Juntando tudo a lista completa dos autovalores é:

$$E_j = \frac{(\hbar\pi)^2}{8ma^2}j^2$$

para $j = 0, 1, 2, \dots$.

Para passar a uma caixa tridimensional, veremos um caso particular de potencial $V(x_1, x_2, x_3)$ dado por $V(x_1, x_2, x_3) = V_1(x_1) + V_2(x_2) + V_3(x_3)$, com cada V_j , $j = 1, 2, 3$ função unidimensional semelhante ao caso anterior.

A caixa pode ser suposta cúbica após uma mudança linear de coordenadas. Então a equação

$$-\frac{\hbar^2}{2m}\nabla^2 u_E + V(x_1, x_2, x_3)u_E = E u_E$$

pode ser resolvida por $u_E(x_1, x_2, x_3) = u_1(x_1)u_2(x_2)u_3(x_3)$, com

$$\left[-\frac{\hbar^2}{2m} \frac{d^2}{dx_j^2} + V_j(x_j) \right] u_j(x_j) = e_j u_j(x_j)$$

com $j = 1, 2, 3$ e $E = e_1 + e_2 + e_3$.

Assim pode-se dizer que ([4]), as autofunções de $(-\frac{\hbar^2}{2m}\nabla^2 + V)$ são os produtos de três autofunções unidimensionais, e os autovalores são a soma dos autovalores unidimensionais correspondentes. Temos que a lista dos autovalores é o conjunto

$$\frac{\hbar^2\pi^2}{8ma^2}(j_1^2 + j_2^2 + j_3^2), \quad j_1, j_2, j_3 \in \mathbb{N}. \quad (6)$$

3.3 Contagem dos níveis de energia

Para compreender a teoria da Termodinâmica Estatística é importante calcular o número aproximado de níveis de energia abaixo de certo valor de energia E . Esse número é a quantidade de pontos do primeiro octante com coordenadas inteiras de norma menor que

$$R = \sqrt{\frac{8ma^2}{\hbar^2\pi^2} E} ,$$

equivalente ao raio da caixa em questão. Esse número pode ser aproximado por um oitavo do volume da esfera:

$$V_o = \frac{1}{8} \frac{4}{3} \pi \left(\frac{8ma^2}{\hbar^2\pi^2} \right)^{\frac{3}{2}} E^{\frac{3}{2}} .$$

Aqui reteremos apenas a informação de que o número de níveis de energia abaixo do valor E é proporcional a $E^{\frac{3}{2}}$.

O caso de N partículas pode ser tratado da seguinte forma. Considera-se a função $\psi(x^{(1)}, x^{(2)}, x^{(3)}, \dots, x^{(N)}, t)$, onde $|\psi|^2$ dá a distribuição de probabilidade de se encontrar a partícula 1 em $x^{(1)}$, a partícula 2 em $x^{(2)}$, etc..., como se fossem $3N$ coordenadas espaciais e uma temporal. Se forem N partículas numa caixa que não interagem entre si, o potencial V poderá ser escrito como

$$V(x^{(1)}, x^{(2)}, x^{(3)}, \dots, x^{(N)}) = \sum_{j=1}^N V(x_1^{(j)}) + V(x_2^{(j)}) + V(x_3^{(j)}) + \dots + V(x_n^{(j)}) .$$

Os níveis de energia globais serão dados pela soma dos níveis de energia das partículas.

Para contar o número de estados com menos energia que E estimaremos o volume da esfera de raio $R = \sqrt{\frac{8ma^2}{\hbar^2\pi^2}E}$ no espaço $3N$ dimensional. Esse volume vale

$$\frac{\pi^{\frac{3N}{2}}}{\Gamma(\frac{3N}{2} + 1)} R^{3N},$$

de forma que o número procurado é da ordem de

$$\left(\frac{8ma^2}{\hbar^2\pi^2}\right)^{\frac{3N}{2}} \frac{\pi^{\frac{3N}{2}}}{\Gamma(\frac{3N}{2} + 1)} E^{\frac{3N}{2}},$$

ou seja, da ordem de $E^{\frac{3N}{2}}$, para N fixo. Não é razoável supor que as partículas não interagem entre si, porém neste caso a resolução da equação de Schrödinger pode se tornar um problema bem mais complicado. Mesmo que haja interação, esse potencial é um bom ponto de partida para os pressupostos estatísticos que servem de base à Termodinâmica Estatística.

3.4 Pressupostos estatísticos da Termodinâmica

A Termodinâmica Estatística ([2], [14]) parte do pressuposto de que a listagem dos níveis de energia num sistema de N partículas admite uma quantidade proporcional a N de “números quânticos”, que são os índices dessa listagem. Por exemplo, uma partícula numa caixa tem 3 números quânticos, j_1, j_2, j_3 , para cada dimensão espacial e as N partículas numa caixa que não interagem têm $3N$ números quânticos, um para cada partícula e dimensão espacial.

Seja $\Omega(E)$ a função que dá o número de estados quânticos do sistema compatíveis com a energia total E . Esta é a principal função da Termodinâmica Estatística. No

entanto, algum cuidado se requer para entendê-la corretamente. Em primeiro lugar, ela pode depender de várias variáveis, ditas externas (por exemplo, o volume da caixa que contém um gás ou um campo magnético, elétrico ou gravitacional externo). Mas aqui, por enquanto, supomos que essas quantidades são fixas.

Em segundo lugar, a energia total só pode assumir uma quantidade enumerável de valores, então não faz sentido fixar um tal E e perguntar por $\Omega(E)$, pois na maioria das vezes a resposta será zero.

A dificuldade é sobrepassada da seguinte forma. Escolhe-se um valor δE pequeno do ponto de vista das medidas macroscópicas mas grande se comparado às diferenças entre níveis de energia de um ou poucos números quânticos. Então define-se $\Omega(E)$ como o número de estados com energia entre $E - \delta E$ e $E + \delta E$.

Veja que se δE for pequeno é razoável supor que $\Omega(E) \simeq \omega(E)\delta E$ para alguma função $\omega(E)$ (uma densidade do número de estados). Se escolhêssemos um valor diferente $\bar{\delta E}$ obteríamos:

$$\bar{\Omega}(E) \simeq \omega(E)\bar{\delta E} = \omega(E)\delta E \frac{\bar{\delta E}}{\delta E} = \Omega(E) \frac{\bar{\delta E}}{\delta E},$$

isto é, a contagem de estados diferiria apenas por um fator constante (que poderia ser absorvido mais adiante por uma constante). A hipótese mais básica da Termodinâmica Estatística é a de que *num sistema isolado com energia total E , todos os estados compatíveis com essa energia são igualmente prováveis*.

Para sermos mais precisos na definição admitimos que “ter energia total E ” corresponde a ter energia total entre $E - \delta E$ e $E + \delta E$. Este princípio tem conseqüências importantes, como veremos mais adiante.

O exemplo das N partículas numa caixa mostra que é razoável supor que $\Omega(E)$ é uma função que cresce muito rapidamente com E .

Observe que $\Omega(E)$ é aproximadamente a derivada da função dada na expressão

$$\left(\frac{8ma^2}{\hbar^2\pi^2}\right)^{\frac{3N}{2}} \frac{\pi^{\frac{3N}{2}}}{\Gamma(\frac{3N}{2} + 1)} E^{\frac{3N}{2}}.$$

multiplicado por $2\delta E$, portanto uma função do tipo $CE^{\frac{3N}{2}-1}$, que também tem crescimento muito rápido com E .

3.5 Interação entre sistemas macroscópicos

Suponha que dois sistemas σ e σ' estejam em contato de tal forma que possam trocar energia mas suas variáveis externas e a energia total não se alterem. Diz-se que esses sistemas têm interação (puramente) térmica.

O calor trocado entre os dois sistemas é a diferença entre a energia inicial (E_i) e a energia final (E_f) de um dos sistemas, contado positivamente. O calor absorvido pelo sistema σ é $\Delta Q = \Delta E = E_f - E_i$, que pode ser negativo se, de fato, o sistema perder energia. O calor absorvido por σ' é $\Delta Q' = \Delta E' = E'_f - E'_i$.

A soma dos calores absorvidos pelos dois sistemas é zero porque a energia total se

conserva ($\Delta Q + \Delta Q' = 0$).

Outra forma de dois sistemas trocarem energia é através de interações mecânicas realizadas através de mudanças nas variáveis externas (tipicamente o volume). Por exemplo um gás numa caixa termicamente isolada, sujeito ao campo gravitacional, quando alçado a grande altura, ou dois gases com uma parede móvel e isolante entre eles.

O trabalho efetuado por um sistema é denotado por ΔW e o outro por $\Delta W'$, e se a interação for puramente mecânica então $\Delta W + \Delta W' = 0$, por causa da conservação de energia. Em geral, porém, as duas formas de trocar energia aparecem juntas. Então $\Delta E = \Delta Q - \Delta W$. Em outras palavras, *o calor absorvido pelo sistema é a diferença de energia do estado inicial para o estado final mais o trabalho realizado pelo sistema*. Na verdade, essa expressão pode ser usada para definir o calor absorvido no caso mais geral onde aparecem interações mecânicas.

Uma situação comum é quando a única variável externa relevante de um sistema é seu volume. Por exemplo, um cilindro com um pistão móvel. A força exercida pelo sistema sobre o pistão é pA , onde p é a pressão e A a área da superfície do pistão. O trabalho infinitesimal realizado pelo sistema é dado por $dW = pAds = pdV$ onde ds é o deslocamento infinitesimal do pistão e dV é a mudança infinitesimal no volume. De fato, a fórmula $dW = pdV$ faz sentido em situações mais gerais do que a de um pistão cilíndrico.

Neste sistema e em outros a energia total é determinada como função da pressão e do

volume através da função $E(V, p)$ (assim como a energia e o volume determinam a pressão), se não houver troca de calor.

Pela observação acima temos

$$p = -\frac{\partial E}{\partial V}.$$

3.6 Formas diferenciais

Vale a pena entender os conceitos apresentados sob o ponto de vista da teoria de Formas Diferenciais ([8]). Falaremos apenas de formas diferenciais definidas em subconjuntos de \mathbb{R}^2 e para conectar mais rapidamente com a teoria usaremos V e p como coordenadas do plano (em vez de x e y , respectivamente).

Uma *1-forma (ou apenas forma) diferencial* definida num aberto $U \subset \mathbb{R}^2$ é uma função contínua que a cada ponto (V, p) de U associa uma transformação linear $\omega(V, p) : \mathbb{R}^2 \rightarrow \mathbb{R}$. Por exemplo, denomina-se dp a forma constante que age como

$$dp(V, p) \cdot (v_1, v_2) = v_2,$$

isto é, que toma a segunda coordenada do vetor em que se aplica. Analogamente, dV toma a primeira coordenada do vetor em que se aplica. Outro exemplo é quando se dá uma função de classe C^1 qualquer $f : U \rightarrow \mathbb{R}$, para a qual se associa a forma diferencial $df = \frac{\partial f}{\partial V}dV + \frac{\partial f}{\partial p}dp$. Isto motiva a seguinte definição.

Definição. Uma forma diferencial ω é dita **exata** se existe $f : U \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ tal que

$\omega = df$. Neste caso, dizemos que a função f é uma **primitiva** da forma ω .

Formas diferenciais podem ser integradas ao longo de caminhos, de acordo com a seguinte definição.

Definição. Dada uma forma diferencial ω e um caminho diferenciável $\lambda : [a, b] \rightarrow U$, definimos a integral de ω ao longo do caminho λ por $\int_{\lambda} \omega = \int_a^b \omega(\lambda(t)) \cdot \lambda'(t) dt$.

Teorema 1. As afirmações são equivalentes:

1. ω é exata.
2. $\int_{\lambda} \omega$ com $\lambda : [a, b] \rightarrow U$ depende apenas das extremidades $\lambda(a)$ e $\lambda(b)$.
3. $\int_{\lambda} \omega = 0$ se λ é um caminho fechado.

No caso em que a energia E é função de V e p podemos definir a forma exata

$$dE = \frac{\partial E}{\partial p} dp + \frac{\partial E}{\partial V} dV = \frac{\partial E}{\partial p} dp - pdV ,$$

onde aproveitamos o fato de que $p = -\frac{\partial E}{\partial V}$.

Porém devemos tomar muito cuidado, pois nem toda forma é exata! Por exemplo, tomemos a forma pdV , que pega a primeira coordenada do vetor e multiplica pelo valor da pressão no ponto. A utilidade dessa forma é que se o sistema vai de (V_i, P_i) a (V_f, P_f) através de uma curva γ então o trabalho realizado é $\int_{\gamma} dW = \int_{\gamma} pdV$.

Se $\gamma(t) = (V(t), p(t))$ temos $\gamma'(t) = (V'(t), p'(t))$ e então $\int_{\gamma} pdV = \int_{t_i}^{t_f} p(t)V'(t)dt$. Se γ for gráfico de uma função $p(V)$ então essa integral é a área sob o gráfico. Mas se o resultado da integral é a área sob o gráfico então o resultado depende do caminho, mostrando que pdV não pode ser uma forma exata!

Pela Seção 3.5, $dE = dQ - dW$, isto é $dQ = dE + dW$ ou $dW = dQ - dE$. Então dQ não pode ser forma exata, pois se fosse $dQ - dE$ seria e então dW seria, contradizendo o que acabamos de ver acima!

Além disso, dW só é forma exata quando o sistema estiver termicamente isolado, que é estabelecido na 1ª Lei da Termodinâmica: “*Se um sistema termicamente isolado é levado de um estado inicial para um final, o trabalho realizado é independente do processo usado*”.

3.7 Entropia e temperatura

Sejam dois sistemas σ_1 e σ_2 termicamente isolados entre si e isolados no seu conjunto, com energia E_1^0 e E_2^0 e com parâmetros externos pré-fixados. O que acontece se removermos a barreira de isolamento térmico (mantendo o conjunto $\sigma_1 \cup \sigma_2$ isolado)? Para responder a esta pergunta, convém olhar as funções $\Omega_1(E_1)$ e $\Omega_2(E_2)$ (Ω_1 e Ω_2 são em geral funções diferentes, pois os sistemas podem ser diferentes).

Antes ou após a remoção, a energia total é constante e igual a $E^0 = E_1^0 + E_2^0$. Mas após a remoção, os sistemas estão livres para trocar energia (calor), desde que a soma permaneça constante.

Agora supomos que após a remoção os dois sistemas atingem o equilíbrio um com respeito ao outro. Como prever a energia total de cada um deles?

Seja Ω o número total de estados acessíveis ao conjunto dos dois sistemas. Então

$$\Omega = \sum_{0 \leq E \leq E^0} \Omega_1(E) \Omega_2(E^0 - E)$$

Chamemos de $\Omega(E) = \Omega_1(E) \Omega_2(E^0 - E)$ o número de estados do sistema com a imposição de que σ_1 esteja com energia E . Então definimos $P(E) = \frac{\Omega(E)}{\Omega}$, a probabilidade de que σ_1 tenha energia E .

Como $\Omega_1(E)$ cresce rápido com E e $\Omega_2(E^0 - E)$ decresce rápido com E então $P(E)$ tem o gráfico como ilustrado na Figura 9.

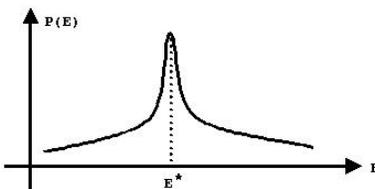


Figura 9: Gráfico de $P(E)$, onde E é a energia de um sistema σ_1 que é subsistema de $\sigma_1 \cup \sigma_2$ e tal que a energia total é fixa, e P é a fração de estados com energia E em comparação com todos os estados possíveis do sistema.

O gráfico ilustra a existência de um ponto de máximo E^* para o qual há maior probabilidade. De fato, o ponto de máximo da função $P(E)$ coincide com o ponto de máximo da função $\ln P(E)$, dada por

$$\ln P(E) = -\ln \Omega + \ln \Omega_1(E) + \ln \Omega_2(E^0 - E) . \quad (7)$$

Então podemos achar E^* simplesmente derivando essa função, ou seja como solução da equação

$$\frac{\partial \ln \Omega_1}{\partial E}(E^*) + (-1) \frac{\partial \ln \Omega_2}{\partial E}(E^0 - E^*) = 0 .$$

Definindo

$$\beta_j = \frac{\partial \ln \Omega_j}{\partial E} ,$$

com $j = 1, 2$, temos

$$\beta_1(E^*) = \beta_2(E^0 - E^*) .$$

As quantidades de que estamos tratando têm nome e significado físico. Por exemplo, $S \equiv k \ln \Omega$, onde k é uma constante positiva, é chamada de *entropia*. Estamos considerando S como uma função da energia E do sistema, portanto S também é uma função de estado.

Observe que a condição de máximo para $\ln P(E)$ é o mesmo que a condição de máximo para a soma das entropias dos dois subsistemas (Equação (7)), ou seja, *as hipóteses probabilísticas implicam que o equilíbrio é atingido quando a soma das entropias é maximizada*.

A quantidade definida pelo símbolo β , que é a derivada da entropia em relação à energia, pode ser suposta uma função monótona da energia (basendo-se nos casos típicos como o dos gases ideais), de forma que E^* fique definida de forma unívoca. Assim, β mede uma propriedade intrínseca dos sistemas e a igualdade $\beta_1 = \beta_2$ acaba por ser uma condição para o equilíbrio, na hipótese de que o sistema passe a maior parte do tempo entre o conjunto de estados mais provável.

Observe que se inicialmente o sistema não atingiu o equilíbrio e a energia de σ_1 é E , com $E < E^*$, então a derivada de $\ln P(E)$ é positiva, o que implica $\beta_1 > \beta_2$. Por outro lado, se $E > E^*$ ela é negativa, isto é $\beta_1 < \beta_2$. Assim, a troca de energia é no sentido do sistema que tem menor β para o que tem maior β .

O significado intuitivo de temperatura nos diz que o calor é trocado do mais quente (temperatura mais alta) para o menos quente (temperatura mais baixa). Definindo a temperatura como um número proporcional ao inverso de β obtemos uma quantificação coerente com a intuição. Assim, a temperatura é definida por $\beta \equiv \frac{1}{kT}$ e pode ser escrita também como $\frac{1}{T} = k\beta = \frac{\partial S}{\partial E}$, com unidade de medida denominada *Kelvin*. A constante k pode ser ajustada para que a diferença entre o ponto de congelamento e o ponto de ebulição da água seja igual a 100 Kelvin (por razões históricas, para que um grau Kelvin se iguale a um grau Celsius, porém de forma a que a escala Kelvin seja uma medida absoluta da temperatura, no sentido que expusemos acima). Isso dá $k = k_B = 1.3807 \times 10^{-23}$ Joules/Kelvin, que é chamada de *constante de Boltzman*.

3.8 Fator integrante

Se um sistema recebe um acréscimo de energia infinitesimal dQ na forma de calor podemos escrever

$$\ln \Omega(E + dQ) - \ln \Omega(E) = \frac{\partial \ln \Omega}{\partial E} dQ = \beta dQ ,$$

isto é,

$$dS = \frac{dQ}{T}.$$

Pode-se argumentar a partir dos pressupostos estatísticos ([14]) que o número de estados acessíveis ao sistema não se altera por processos quasi-estáticos (isto é, lentos a ponto de permitir ao sistema permanecer em equilíbrio durante o processo) envolvendo apenas trabalho mecânico. Em outras palavras, num sistema termicamente isolado (ou seja, para o qual não se permite troca de calor) onde seus parâmetros externos (tipicamente o volume) são alterados de forma quasi-estática não há mudança na entropia. Assim, a igualdade $dS = \frac{dQ}{T}$ vale mesmo quando há interações térmica e mecânica simultaneamente.

Observe que embora dQ não seja uma diferencial exata ela passa a ser quando dividida pela temperatura, pois dS é exata. Assim o inverso da temperatura é um *fator integrante* que multiplica a forma dQ e a torna exata.

3.9 A função entropia e a relação de Euler

Em resumo, a teoria da Termodinâmica Estatística pressupõe que o sistema tenha uma *função entropia*

$$S = S(E, V, N_1, \dots, N_r)$$

onde E é a energia, V é o volume e N_i é o número de moléculas do tipo B_i , $i = 1, \dots, r$.

Essa função deve satisfazer algumas propriedades, que podem ser estabelecidas de modo

axiomático ou sugeridas de argumentos estatísticos, como fizemos acima, desde que se defina a entropia como proporcional ao logaritmo do número de estados acessíveis ao sistema.

Em primeiro lugar, a entropia é uma função estritamente crescente da energia. De fato, como $\frac{\partial S}{\partial E} = \beta$ é tanto menor quanto maior for a energia, supõe-se que $\frac{\partial^2 S}{\partial E^2} < 0$ (admite-se boas propriedades de diferenciabilidade para que essas expressões façam sentido).

As variáveis das quais depende a entropia são chamadas *extensivas*. Pressupõe-se que o equilíbrio macroscópico do sistema seja totalmente caracterizado por essas variáveis.

Os jargões que dizem “a entropia só cresce” ou “a entropia é sempre maximizada” têm formulação precisa. Um sistema composto (σ_1 com σ_2) pode ser simplesmente a junção “lado a lado” dos dois sistemas, sem que haja interação entre eles, de qualquer espécie. Para não haver interação, é preciso que haja várias barreiras: *térmica*, para não haver troca de energia na forma de calor; *mecânica*, para não haver alterações de volume; ou *química*, para não haver transferências de moléculas de um sistema para o outro. No entanto, se uma das barreiras for removida, as variáveis externas que podem se alterar passarão, dentro das suas possibilidades, a assumir valores que maximizam a entropia. Por exemplo, se a barreira térmica for removida as energias E^1 e E^2 dos dois sistemas assumirão valores tais que a entropia total

$$S_1(E^{(1)}, V^{(1)}, N_1^{(1)}, \dots, N_r^{(1)}) + S_2(E^{(2)}, V^{(2)}, N_1^{(2)}, \dots, N_r^{(2)}),$$

seja maximizada, respeitando-se que a soma $E^{(1)} + E^{(2)}$ permaneça constante e igual à

energia total do sistema composto.

Assume-se também que a entropia seja aditiva sobre suas partes constituintes, o que implica, de fato, que a entropia é uma função homogênea (de primeira ordem) das suas variáveis. Isto é, dado $\alpha > 0$ então

$$S(\alpha E, \alpha V, \alpha N_1, \dots, \alpha N_r) = \alpha S(E, V, N_1, \dots, N_r) .$$

O fato de que S é uma função monótona da variável E implica que esta variável pode ser isolada utilizando S como variável:

$$E = E(S, V, N_1, \dots, N_r) .$$

Assim, a Termodinâmica pode ser apresentada de forma alternativa. Essa função também é homogênea de primeira ordem. Derivando a relação de homogeneidade

$$E(\alpha S, \alpha V, \alpha N_1, \dots, \alpha N_r) = \alpha E(S, V, N_1, \dots, N_r) .$$

em relação a α e tomando $\alpha = 1$ segue que

$$E(S, V, N_1, \dots, N_r) = \frac{\partial E}{\partial S} S + \frac{\partial E}{\partial V} V + \frac{\partial E}{\partial N_1} N_1 + \dots + \frac{\partial E}{\partial N_r} N_r .$$

Como vimos anteriormente, $-\frac{\partial E}{\partial V}$ é a pressão p e $\frac{\partial E}{\partial S}$ é o inverso de $\frac{\partial S}{\partial E}$, isto é, a temperatura T . As derivadas parciais $\frac{\partial E}{\partial N_i}$ recebem a denominação de μ_i e são chamadas de *potenciais químicos*. Então, resumidamente,

$$E = TS - pV + \mu_1 N_1 + \dots + \mu_r N_r ,$$

equação que é conhecida como *relação de Euler*. Deve-se atentar para o fato de que as variáveis T , p , μ_i (ditas *intensivas*) são, nesta equação, funções de S , V , N_i .

Vimos que o princípio de maximização da entropia implica que a remoção da barreira térmica leva à igualdade de temperaturas (ver Seção 3.7). Se também for removida a barreira volumétrica o mesmo princípio implicará na igualdade de temperaturas e pressões. Então temperatura e pressão podem ser vistas como *potenciais*, pois indicam o sentido e a intensidade das mudanças na energia e no volume. Os potenciais químicos têm papel semelhante (daí seu nome): a remoção das barreiras térmica e química proporciona a igualdade de temperaturas e potenciais químicos.

Também é possível formular o princípio de maximização da entropia como um princípio de minimização da energia ([2] ou [14]). Se o equilíbrio entre dois sistemas sem barreira térmica é aquele dado por E^1 e E^2 , tais que $E^1 + E^2 = E_T$ (onde E_T é a energia total fixa) e que maximizam a entropia $S^1 + S^2$, se chamarmos de S_T o valor de $S^1 + S^2$ no equilíbrio, então valerá também que S^1 e S^2 são dados pela condição de que $S^1 + S^2 = S_T$ e de tal maneira a minimizar $E^1 + E^2$.

3.10 A energia livre de Gibbs

Suponha agora que tenhamos dois sistemas σ e σ' , onde σ' é um *reservatório de temperatura e pressão*. Isto significa que σ e σ' podem trocar calor e trabalho mecânico por pressão, mas σ' é tão grande que sua temperatura T_0 e sua pressão p_0 não se alteram. Assim, o

sistema σ terá sempre, no equilíbrio, as mesmas temperatura e pressão que o reservatório.

Este tipo de hipótese condiz com as condições encontradas em laboratório, especialmente quando se realizam reações químicas. A atmosfera é um reservatório de pressão e temperatura, condições nas quais se dão as reações. Durante o processo de reação pode haver alterações ΔS e $\Delta S'$ nas entropias de σ e σ' , mas sempre no sentido de aumentar a entropia, isto é a variação da entropia total $\Delta S_T = \Delta S + \Delta S'$ é sempre não negativa.

Seja ΔQ o calor transferido de σ' para σ (com sinal). Então, como a temperatura do reservatório não se altera durante o processo,

$$\Delta S' = -\frac{\Delta Q}{T_0},$$

por causa da relação de Euler, dada na Seção 3.9. Mas, pelo que vimos na Seção 3.5, temos

$$\Delta Q = \Delta E + p_0 \Delta V,$$

onde ΔE é a mudança de energia em σ e $p_0 \Delta V$ é o trabalho feito em σ contra a pressão constante em σ' , supondo que este seja o único tipo de interação mecânica entre os sistemas.

Então

$$\Delta S_T = \Delta S - \frac{\Delta Q}{T_0} = \frac{1}{T_0} [T_0 \Delta S - \Delta E - p_0 \Delta V].$$

Denominando

$$G_0 = E + p_0 V - T_0 S$$

teremos

$$\Delta S_T = -\frac{\Delta G_0}{T_0} .$$

Em palavras, dG_0 representa o acréscimo infinitesimal (com sinal) de energia mais trabalho feito contra a pressão do reservatório menos a energia recebida em forma de calor, do reservatório, ou seja, trata-se da energia efetiva gerada pelo processo.

Como $\Delta S_T \geq 0$ isso implica que $\Delta G_0 \leq 0$, logo a situação de equilíbrio é caracterizada pela condição de que a quantidade G_0 seja mínima.

Dá-se o nome de *energia livre de Gibbs* à grandeza G_0 .

Mais formalmente define-se a energia livre de Gibbs de um sistema como a função

$$G = G(T, p, N_1, \dots, N_r)$$

obtida de $E = E(S, V, N_1, \dots, N_r)$ da seguinte forma: toma-se o sistema

$$\begin{cases} T = \frac{\partial E}{\partial S}(S, V, N_1, \dots, N_r) \\ -p = \frac{\partial E}{\partial V}(S, V, N_1, \dots, N_r) \end{cases} .$$

resolvendo S e V em função de p e T e escrevendo $G = E - TS + pV$ em função de p, T, N_1, \dots, N_r . A função G é uma *transformação de Legendre* da função E e ambas contêm todas as informações sobre o sistema.

3.11 Equilíbrio químico

Uma equação química pode ser escrita na forma

$$\sum_{i=1}^r \nu_i B_i = 0 ,$$

onde os ν_i , chamados de *coeficientes estequiométricos*, são números inteiros. O sinal de cada ν_i indica “o lado” da reação em que está o composto B_i ao qual se refere. Por exemplo, na reação



podemos chamar $B_1 = H_2$, $B_2 = O_2$ e $B_3 = H_2O$, com $\nu_1 = -2$, $\nu_2 = -1$ e $\nu_3 = +2$, convencionando-se que os sinais positivos correspondam ao lado direito da equação (produtos) e os sinais negativos ao lado esquerdo da equação (reagentes).

Como sempre, seja N_i o número de moléculas de B_i presentes no sistema. Supondo que os compostos sejam colocados juntos, se houver reação química então os números N_i poderão se alterar, mas de uma maneira que preservem o número de átomos, isto é, respeitando a equação de reação. Ou seja, $\Delta N_i = \lambda \nu_i$, para um certo λ , que representa a “quantidade de reação” efetuada.

Na situação de equilíbrio em que o sistema esteja em contato com um reservatório de temperatura T_0 e pressão p_0 , a energia livre de Gibbs é a mínima, isto é, $dG_0 = 0$. Então

$$dG_0 = d(E - T_0S + p_0V) = \sum_i \mu_i dN_i = 0 .$$

Substituindo dN_i por $\lambda\nu_i$, temos

$$\sum_i \mu_i \nu_i = 0 , \quad (8)$$

que diz em que condições podemos achar o equilíbrio, desde que conheçamos como se comportam os potenciais químicos.

Em situações ideais onde os compostos são gases é possível deduzir as funções de potencial químico a partir dos pressupostos quânticos e estatísticos. Como

$$\mu_i = \mu_i(T, p, N_1, \dots, N_r) ,$$

$p = p_0$ e $T = T_0$, e ainda supondo que não haja dependência em relação ao número de outras moléculas, o potencial químico é essencialmente uma função de N_i . No caso de compostos que reagem em um solvente, observa-se experimentalmente ([11],[12],[6]) que

$$\mu_i = \mu_i^0 + k_B T \ln[B_i] ,$$

onde $[B_i]$ é a concentração da molécula B_i em Molar, ou seja, μ_i^0 é o potencial químico para a concentração de 1 Molar. Substituindo na equação (8), temos

$$\sum_{i=1}^r \nu_i (\mu_i^0 + k_B T \ln[B_i]) = 0 .$$

Podemos também multiplicar esta equação por N_A , o *número de Avogadro*, que representa o número de moléculas que formam 1 mol. A constante $N_A k_B$ é denotada por R e é

chamada de “constante universal dos gases”. Assim,

$$\sum_{i=1}^r \nu_i \mu_i^0 N_A + RT \sum_{i=1}^r \nu_i \ln[B_i] = 0 .$$

Se chamarmos de ΔG^0 a soma $\sum_{i=1}^r \nu_i \mu_i^0 N_A$ (justificaremos isso logo em seguida) e esse valor for conhecido, então esta é a equação sobre as concentrações do equilíbrio, e pode ser reformulada para

$$\prod_{i=1}^r [B_i]^{\nu_i} = e^{-\frac{\Delta G^0}{RT}} = K_{eq} , \quad (9)$$

que é conhecida como *equação do equilíbrio químico*.

Para entender melhor o significado do ΔG^0 , primeiro definiremos o que é “1 mol de reação”. Se s é tal que

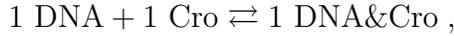
$$\nu_1, \nu_2, \dots, \nu_s < 0 < \nu_{s+1}, \nu_{s+2}, \dots, \nu_r ,$$

isto é, os primeiros s coeficientes estequiométricos se referem aos reagentes e os demais aos produtos, e além disso os coeficientes não têm fator comum, então dizemos que ocorreu “1 mol de reação”, se $N_A \nu_1$ moléculas de B_1 , $N_A \nu_2$ moléculas de B_2 , ..., $N_A \nu_s$ moléculas de B_s se transformaram em $N_A \nu_{s+1}$ moléculas de B_{s+1} , $N_A \nu_{s+2}$ moléculas de B_{s+2} , ..., $N_A \nu_r$ moléculas de B_r .

Como $\mu_i^0 N_A$ representa a energia livre de Gibbs de um mol da substância B_i à concentração de 1 Molar, então $\Delta G^0 = \sum_{i=1}^r \nu_i \mu_i^0 N_A$ é a diferença entre a soma das energias livres de ν_1 moles de B_1 , ν_2 moles de B_2 , ..., ν_s moles de B_s e a soma das energias livres de

ν_{s+1} moles de B_{s+1} , ν_{s+2} moles de B_{s+2} , ..., ν_r moles de B_r .

Esse dado ΔG^0 é mensurável. Por exemplo, no modelo temos uma reação do tipo



onde a energia da parte direita da equação é menor que a da parte esquerda, fazendo com que ΔG^0 seja menor que zero.

A equação de equilíbrio químico é comumente apresentada na forma

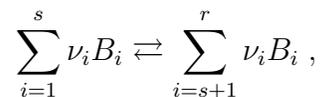
$$\frac{[B_{s+1}]^{\nu_{s+1}} \dots [B_r]^{\nu_r}}{[B_1]^{\nu_1} \dots [B_s]^{\nu_s}} = K_{eq} = e^{-\frac{\Delta G^0}{RT}} . \quad (10)$$

Essa equação será usada no próximo Capítulo para se obter as probabilidades (forças) de cada configuração sobre os sítios operadores em função dos valores de energia livre e das concentrações. Essas probabilidades serão usadas na construção da equação diferencial.

É importante atentar para o fato de que **a equação não é invariante por uma mudança de unidades de medida de concentração**: se elas forem mudadas, muda a constante de equilíbrio. Mas a própria definição de ΔG^0 depende da unidade escolhida para a concentração.

3.12 Dedução probabilística da equação de equilíbrio

Considerando uma equação na forma



sabe-se que a probabilidade de ocorrência da reação da esquerda para a direita, que chamaremos aqui de P_+ , deve ser proporcional à probabilidade de se encontrar simultaneamente, num determinado elemento de volume, ν_1 moléculas de B_1 , \dots , ν_s moléculas de B_s . Supondo que as probabilidades são independentes, então a probabilidade de se encontrar uma molécula de B_i será proporcional a $[B_i]$. Portanto $P_+ = K_+ \prod_{i=1}^s [B_i]^{\nu_i}$, onde K_+ é uma constante de proporcionalidade.

Analogamente, definindo P_- como a probabilidade de ocorrência da reação da direita para a esquerda, teremos $P_- = K_- \prod_{i=s+1}^r [B_i]^{\nu_i}$

Numa situação de equilíbrio deveremos ter $P_+ = P_-$ e conseqüentemente

$$\frac{\prod_{i=s+1}^r [B_i]^{\nu_i}}{\prod_{i=1}^s [B_i]^{\nu_i}} = \frac{K_+}{K_-} = K_{eq} ,$$

ou seja, pode-se chegar à Equação 10 por argumentos probabilísticos, mas esse argumento não mostra como obter a constante de equilíbrio de forma independente.

4 O modelo

4.1 Probabilidades das configurações

Com o intuito de chegar ao nosso modelo, interessa-nos analisar o equilíbrio de uma equação do tipo



onde um número finito de reagentes formam um único produto e para a qual vale

$$\frac{[C]^{\nu_C}}{[B_1]^{\nu_1} \dots [B_r]^{\nu_r}} = K_{eq} = e^{-\frac{\Delta G^0}{RT}} ; \quad (11)$$

onde K_{eq} é a constante de equilíbrio da equação, ΔG^0 é a energia livre de Gibbs associada à equação, R é a constante universal dos gases e T é a temperatura absoluta no equilíbrio. Vimos no Capítulo 3 como se deduz essa equação a partir da teoria da Termodinâmica Estatística.

Esse tipo de equação química é aqui utilizado porque sempre teremos várias moléculas de proteínas que podem se ligar a apenas uma molécula de DNA, gerando uma configuração caracterizada pela molécula de DNA com uma ou mais destas proteínas ligadas a ela. A cada configuração que considerarmos, o que irá variar serão os tipos de proteínas que se ligarão ao DNA e a quantidade de cada um destes tipos que se ligará ao mesmo tempo.

Mais precisamente, teremos



onde Conf é uma das configurações de equilíbrio possíveis. Na Tabela 6 listaremos todas as configurações possíveis e os dados correspondentes.

Assim, temos

$$\frac{[\text{Conf}]}{[\text{DNA}]} = [\text{Rep}]^{\nu_{\text{Rep}}} [\text{Cro}]^{\nu_{\text{Cro}}} [\text{RNAP}]^{\nu_{\text{RNAP}}} K, \quad (12)$$

onde K é a constante de equilíbrio da reação. Esta informação vai nos ajudar a deduzir o modelo.

Como já vimos pela Equação (10), a constante K se relaciona com a energia livre de Gibbs da configuração onde as proteínas e a RNAP estão ligadas à molécula de DNA, em quantidades determinadas por seus coeficientes ν_{Rep} , ν_{Cro} e ν_{RNAP} . Na Tabela 6 são fornecidos os valores de energia livre de Gibbs e não diretamente as constantes de equilíbrio.

A rigor, o quociente

$$\frac{[\text{Conf}]}{[\text{DNA}]}$$

faz pouco sentido, visto que o número de moléculas de DNA e Conf (situação em que a molécula de DNA está ligada a uma determinada quantidade de proteínas) só pode ser 0 ou 1 e, ainda, quando um é 0 o outro é 1, e vice-versa.

Mas se levarmos em consideração que o equilíbrio é estatístico e não estático, o quociente será a proporção de tempo em que o sistema apresenta a configuração de equilíbrio em

relação ao tempo em que não apresenta a configuração. Seria então algo como

$$\frac{\tau_{Conf}}{\tau_{DNA}},$$

denotando por τ os tempos respectivos.

Como existem várias configurações possíveis no modelo completo, temos que considerar a possibilidade de várias reações simultaneamente e para cada uma delas tem-se a equação

$$\frac{\tau_{Conf_i}}{\tau_{DNA}} = [\text{Rep}]^{\nu_{Rep}^i} [\text{Cro}]^{\nu_{Cro}^i} [\text{RNAP}]^{\nu_{RNAP}^i} K_i. \quad (13)$$

Admitindo que podemos fixar, arbitrariamente, um intervalo de tempo de tamanho τ e que, dentro deste intervalo, cada situação corresponderá a uma fração de τ , teremos

$$\tau = \tau_{DNA} + \sum_{i=1}^s \tau_{Conf_i}.$$

A proporção de tempo da configuração i com relação ao tempo τ é

$$\frac{\tau_{Conf_i}}{\tau} = \frac{\tau_{Conf_i}}{\tau_{DNA} + \sum_i \tau_{Conf_i}}. \quad (14)$$

Observemos que o lado esquerdo da equação nos dá a probabilidade de se encontrar a configuração i e o lado direito pode ser desenvolvido com a Equação (13). Então

$$Prob(Conf_i) = \frac{[\text{Rep}]^{\nu_{Rep}^i} [\text{Cro}]^{\nu_{Cro}^i} [\text{RNAP}]^{\nu_{RNAP}^i} K_i}{1 + \sum_{j=1}^s [\text{Rep}]^{\nu_{Rep}^j} [\text{Cro}]^{\nu_{Cro}^j} [\text{RNAP}]^{\nu_{RNAP}^j} K_j}. \quad (15)$$

Na formulação das equações diferenciais, levaremos em conta o acoplamento das probabilidades referentes às configurações onde ocorre ativação dos promotores P_R e P_{RM} ,

ou seja, configurações que possibilitam a produção das proteínas regulatórias Rep e Cro. Desta forma, estas probabilidades, associadas a outros parâmetros, serão utilizadas para determinar as taxas de produção para essas proteínas.

4.2 Tabela de configurações

Ao descrevermos todas as configurações possíveis para o modelo, utilizaremos várias afirmações de natureza físico-química das interações moleculares, como regras de ligação, interação com cooperatividade e exclusão. Assim, determinaremos as possíveis combinações entre as moléculas de dímeros das proteínas Rep, Cro e RNAP com seus sítios operadores e promotores.

Passaremos agora a relatar as afirmações necessárias para entender as configurações moleculares microscópicas que podem ocorrer nos sítios operadores O_R 's. Observando novamente as Figuras 5, 7 e 8 encontradas no Capítulo 2 tais afirmações podem ser melhor compreendidas. São elas:

- 1. Os 3 sítios operadores de ligação são específicos, não se sobrepõem e em cada um liga-se apenas uma proteína em forma de dímero por vez.*
- 2. Quando há cooperatividade entre as proteínas ao se ligarem na molécula de DNA, o valor da energia livre (em módulo) desta configuração é maior que a soma das energias livres das configurações em que as proteínas estão ligadas separadamente.*

3. Não há interação cooperativa entre um dímero de Cro e qualquer outra proteína, incluindo outro dímero de Cro no sítio adjacente.
4. Nos operadores com 3 dímeros de Rep ligados somente há cooperatividade entre dímeros em O_{R1} e O_{R2} .
5. Se a RNAP liga-se em P_{RM} , previne que os dímeros de Cro ou Rep liguem-se em O_{R3} e vice-versa.
6. Se a RNAP liga-se em P_R , previne que os dímeros de Rep ou Cro liguem-se em O_{R1} e O_{R2} e vice-versa.
7. Duas RNAP podem ocupar P_{RM} e P_R simultaneamente, mas não há interação entre as polimerases nos promotores adjacentes ¹.

As afinidades das configurações nos sítios operadores são determinadas pela energia livre de Gibbs, cujos valores estão expostos na Tabela 6. Esses valores, associados às concentrações de cada proteína, determinarão as probabilidades de cada configuração.

Quanto mais negativa for a variação de energia livre para a ligação, maior é a afinidade entre proteínas e DNA. Algumas configurações têm maior energia livre do que poderia

¹Há informações contraditórias nas nossas referências sobre a possibilidade de transcrição de genes nesta configuração. No artigo [15], observamos que ao mesmo tempo em que foi dito explicitamente que não haveria transcrição de nenhum dos dois, mais adiante os autores supõem implicitamente que ambos os genes podem ser transcritos. Nós utilizaremos a segunda opção.

se supor a partir da soma das energias livres de cada configuração separadamente, como dissemos acima na segunda afirmação e comentamos na Seção 2.10. Por exemplo,

✓ Um dímero da proteína Cro em O_{R2} corresponde a uma variação de energia de -10.8 Kcal/mol, ou seja, $\Delta G'_2 = -10.8$.

✓ Um dímero da proteína Cro em O_{R3} , sozinho, tem $\Delta G'_3 = -12.1$.

✓ Um dímero da proteína Cro em O_{R3} e outro em O_{R2} têm, juntos, $\Delta G'_3 + \Delta G'_2 = (-12.1) + (-10.8) = -22.9$, pois não há cooperatividade.

✓ Um dímero da proteína Rep em O_{R1} corresponde a uma variação de energia de -11,7 Kcal/mol, ou seja, $\Delta G_1 = -11.7$.

✓ Um dímero da proteína Rep em O_{R2} , sozinho, tem $\Delta G_2 = -10.1$.

✓ Um dímero da proteína Rep em O_{R1} e outro em O_{R2} têm, juntos, $\Delta G_1 + \Delta G_2 + \Delta G_{12} = (-11.7) + (-10.1) + (-1.9) = -23.7$, devido à presença de cooperatividade.

Na Tabela 6 a primeira coluna numera as configurações possíveis. O símbolo (t) ao lado do número indica que aquela configuração tem possibilidade de transcrição de pelo menos um dos genes, por causa da presença da RNAP.

A segunda coluna expõe os tipos de proteínas ligadas nos três sítios operadores e promotores. Cada dímero de Rep ou Cro pode se ligar num dos três sítios operadores. Nesta coluna, 0 indica que o sítio está desocupado, R_2 representa um dímero de Rep ligado ao sítio operador indicado pela posição deste símbolo e, analogamente, para um dímero de

s	O_{R3}	O_{R2}	O_{R1}	monômio	$\Delta G(\frac{Kcal}{mol})$
1	0	0	0	$x^0y^0z^0$	0.0
2	0	0	R_2	$x^1y^0z^0$	-11.7
3	0	R_2	0	$x^1y^0z^0$	-10.1
4	R_2	0	0	$x^1y^0z^0$	-10.1
5	0	0	C_2	$x^0y^1z^0$	-10.8
6	0	C_2	0	$x^0y^1z^0$	-10.8
7	C_2	0	0	$x^0y^1z^0$	-12.1
(t)8	RNAp	0	0	$x^0y^0z^1$	-11.5
(t)9	0	RNAp	0	$x^0y^0z^1$	-12.5
10	0	R_2^*	R_2	$x^2y^0z^0$	-23.7
11	R_2	0	R_2	$x^2y^0z^0$	-21.8
12	R_2^*	R_2	0	$x^2y^0z^0$	-22.2
13	0	C_2	C_2	$x^0y^2z^0$	-21.6
14	C_2	0	C_2	$x^0y^2z^0$	-22.9
15	C_2	C_2	0	$x^0y^2z^0$	-22.9
(t)16	RNAp	RNAp	0	$x^0y^0z^2$	-24.0
17	0	C_2	R_2	$x^1y^1z^0$	-22.5
18	0	R_2	C_2	$x^1y^1z^0$	-20.9
19	R_2	0	C_2	$x^1y^1z^0$	-20.9
20	C_2	0	R_2	$x^1y^1z^0$	-23.8
21	R_2	C_2	0	$x^1y^1z^0$	-20.9
22	C_2	R_2	0	$x^1y^1z^0$	-22.2
(t)23	R_2	RNAp	0	$x^1y^0z^1$	-22.6
(t)24	RNAp	R_2	0	$x^1y^0z^1$	-21.6
(t)25	RNAp	0	R_2	$x^1y^0z^1$	-23.2
(t)26	C_2	RNAp	0	$x^0y^1z^1$	-24.6
(t)27	RNAp	C_2	0	$x^0y^1z^1$	-22.3
(t)28	RNAp	0	C_2	$x^0y^1z^1$	-22.3
29	R_2	R_2^*	R_2	$x^3y^0z^0$	-33.8
30	C_2	C_2	C_2	$x^0y^3z^0$	-33.7
31	C_2	R_2^*	R_2	$x^2y^1z^0$	-35.8
32	R_2	C_2	R_2	$x^2y^1z^0$	-32.6
33	R_2^*	R_2	C_2	$x^2y^1z^0$	-33.0
34	R_2	C_2	C_2	$x^1y^2z^0$	-31.7
35	C_2	R_2	C_2	$x^1y^2z^0$	-33.0
36	C_2	C_2	R_2	$x^1y^2z^0$	-34.6
(t)37	RNAp	R_2^*	R_2	$x^2y^0z^1$	-35.2
(t)38	RNAp	C_2	C_2	$x^0y^2z^1$	-33.1
(t)39	RNAp	C_2	R_2	$x^1y^1z^1$	-34.0
(t)40	RNAp	R_2	C_2	$x^1y^1z^1$	-32.4

Tabela 6: Tabela de configurações

Cro utilizamos o símbolo C_2 .

O símbolo RNAP na coluna de O_{R3} indica que a enzima RNAP está ligada no sítio promotor P_{RM} , em posição de transcrever o gene cI , que produz a proteína Rep. Para indicar que a enzima RNAP está ligada ao sítio promotor P_R colocamos o símbolo RNAP na coluna de O_{R2} e deixamos vazia a posição na coluna de O_{R1} .

Quando colocamos o símbolo R_2^* , indicamos que este dímero e o seu adjacente agem com cooperatividade.

Os monômios da terceira coluna refletem a configuração da segunda coluna, pois os expoentes de x , y e z correspondem aos números de moléculas de R_2 , C_2 e RNAP, respectivamente, presentes na configuração. Esses monômios aparecem nos numeradores das probabilidades das configurações, segundo a Equação (10). No próximo Capítulo, usaremos x , y e z para denotar as concentrações de R_2 , C_2 e RNAP, respectivamente, simplificando a notação.

Na quarta coluna temos os valores totais de energia livre de Gibbs relacionadas com a configuração. O estado referencial para toda energia livre é o genoma do fago sem proteínas regulatórias. Esses valores são dados em [15] e foram obtidos através de experiências *in vivo* e *in vitro*.

De modo geral e resumindo, a probabilidade de cada configuração i pode ser escrita

como:

$$Prob[Conf_i] = \frac{e^{-\Delta G_i/RT} x^{j_i} y^{k_i} z^{m_i}}{\sum_{l=1}^{40} e^{-\Delta G_l/RT} x^{j_l} y^{k_l} z^{m_l}} ,$$

onde

- ΔG_i é a energia livre de Gibbs da configuração i , dada em Kcal/mol.
- $R = 0.001987093690$ Kcal/mol Kelvin é a constante universal dos gases.
- $T = 310K$ é a temperatura absoluta dos experimentos.
- x, y, z são as concentrações de R_2, C_2 e da enzima RNA polimerase, dadas em Molar = mol/litro;
- j_i, k_i, m_i são os índices que indicam o número de moléculas de cada tipo ligadas na configuração i (0 - 3).

4.3 Equação diferencial

A expressão que fornece a probabilidade de se encontrar uma determinada configuração s será utilizada na equação diferencial que dará a variação nas concentrações das proteínas, como um valor que indica a possibilidade de transcrição dos genes envolvidos. Porém este número sozinho ainda não é o suficiente para determinar a produção de novas proteínas.

O simples fato de ocorrer uma configuração com possibilidade de transcrição não indica que ela realmente ocorrerá. Como foi dito anteriormente, existe um valor conhecido

como taxa de formação de *complexos abertos* que representa a chance de ocorrer a condição ideal para a transcrição. Desta forma, podemos dizer que a contribuição de cada configuração para as mudanças nas concentrações das proteínas será o produto de três fatores: a probabilidade da configuração, a taxa de isomerização da RNAP em complexos abertos, e o número de moléculas de proteínas sintetizadas por transcrição (dada em unidades de concentração).

As probabilidades que influenciam ativamente nas equações são aquelas referentes às configurações que indicam a possibilidade de transcrição do gene correspondente à proteína referida na equação. Pelas descrições anteriores dos processos biológicos, dadas no Capítulo 2, sabemos que as configurações citadas acima são aquelas em que a RNAP está ligada ao sítio promotor do gene que, ao ser transcrito, permitirá a produção da proteína à qual se refere.

Então, ao escrevermos a equação que descreve a variação na concentração de dímeros de Rep, levamos em conta as probabilidades das configurações em que a RNAP está ligada ao promotor P_{RM} correspondente ao gene *cI*. Analogamente, ao escrevermos a equação que descreve a variação na concentração de dímeros da proteína Cro, levamos em conta as probabilidades das configurações em que a RNAP está ligada ao promotor P_R correspondente ao gene *cro*.

Como foi descrito no Capítulo 2, no caso do gene *cI*, existe uma cooperatividade entre

a RNAP e o dímero de Rep, quando este se encontra no sítio operador O_{R2} . Essa cooperatividade será representada no modelo por um aumento da energia da configuração. Além disso, há ainda um considerável aumento na chance de formação de complexos abertos nestas ligações, por isso teremos taxas distintas a serem consideradas nos casos em que ocorre e em que não ocorre esta cooperatividade.

Podemos então, escrever as equações diferenciais do nosso modelo da seguinte forma:

$$\frac{d[R_2]}{dt} = \left[Prob_1[P_{RM}\&RNAP] \cdot k_{PRM1} + Prob_2[P_{RM}\&RNAP] \cdot k_{PRM2} \right] \cdot A_{R2} - [R_2]k_{dR}$$

e

$$\frac{d[C_2]}{dt} = Prob[P_R\&RNAP] \cdot k_{PR} \cdot A_{C2} - [C_2]k_{dC}$$

onde

- $Prob_1[P_{RM}\&RNAP]$ é um valor entre 0 e 1, pois se trata da soma das probabilidades referentes às configurações onde a RNAP está ligada em P_{RM} e é ajudada pelo dímero de Rep ligado em O_{R2} , ou seja, $Prob_1[P_{RM}\&RNAP] = Prob[Conf_{24}] + Prob[Conf_{37}] + Prob[Conf_{40}]$;
- $Prob_2[P_{RM}\&RNAP]$ é a soma das probabilidades referentes às configurações onde a RNAP está ligada em P_{RM} e sem a ajuda do dímero de Rep ligada em O_{R2} , ou seja, $Prob_2[P_{RM}\&RNAP] = Prob[Conf_8] + Prob[Conf_{16}] + Prob[Conf_{25}] + Prob[Conf_{27}] + Prob[Conf_{28}] + Prob[Conf_{38}] + Prob[Conf_{39}]$;

- $Prob[P_R \& RNAP]$ é a soma das probabilidades referentes às configurações onde a RNAP está ligada em P_R , ou seja, $Prob[P_R \& RNAP] = Prob[Conf_9] + Prob[Conf_{16}] + Prob[Conf_{23}] + Prob[Conf_{26}]$;
- k_{PRM1} é a taxa de isomerização da RNAP no promotor P_{RM} com cooperatividade do dímero de Rep em O_{R2} - dada em OC s^{-1} complexos abertos formados por segundo;
- k_{PRM2} é a taxa de isomerização da RNAP no promotor P_{RM} sem cooperatividade - dada em OC s^{-1} ;
- k_{PR} é a taxa de isomerização da RNAP no promotor P_R - dada em OC s^{-1} ;
- A_{R2} é a quantidade de dímeros de Rep formados em cada transcrição - dada em molar OC^{-1} (vide Seção 4.5 para comentários sobre esta unidade de medida) ;
- A_{C2} é a quantidade de dímeros de Cro formados em cada transcrição - dada em molar OC^{-1} ;
- k_{dR} é a taxa de degradação da proteína Rep - dada em s^{-1} ;
- k_{dC} é a taxa de degradação da proteína Cro - dada em s^{-1} .

A integração numérica das equações diferenciais formuladas produz as concentrações celulares dos dímeros Cro e Rep ao longo do tempo.

4.4 Comentários sobre o modelo original

O objetivo do modelo originalmente formulado no artigo [15] era observar o comportamento das concentrações das proteínas ao longo de uma hora, sob condições iniciais e parâmetros de interesse.

Algumas diferenças podem ser apontadas entre os dois modelos apresentados, tanto na maneira de apresentar as equações como em outros aspectos de natureza qualitativa.

No original a segunda equação dá a variação da concentração de dímeros da proteína Cro em dímeros, tal como expusemos aqui, mas a primeira equação, relativa à proteína Rep, dá essa variação em monômeros. Isso é possível porque, como foi dito no Capítulo 2, há um equilíbrio entre as concentrações de monômeros e dímeros desta proteína presentes na célula, embora somente aquelas que estão formando dímeros participem da regulação gênica. No entanto dentro das expressões das probabilidades foi usada a concentração de dímeros. Fora destas expressões, todas as outras informações, inclusive a própria concentração, foram dadas com respeito a monômeros - a quantidade de proteínas formada por transcrição (A_R), a taxa de degradação (k_{dR}). Isso pode gerar uma certa confusão, por isso no modelo que formulamos ambas as equações já representam a variação nas concentrações das proteínas em forma de dímeros, e portanto todas as informações dizem respeito a este formato de ambas as proteínas.

É claro que, apesar desta diferença, como ambos foram formulados basicamente utili-

zando as mesmas características, podemos dizer que um é uma espécie de adaptação do outro. Para relacioná-los, vamos utilizar $[R]$ para indicar a concentração total de monômeros de Rep (livres ou não) presentes na bactéria e $[R_2]$ para a concentração em dímeros desta mesma proteína. Como foi dito na Seção 2.9 anteriormente, utilizando os dados de [13] temos que $[R_2] = 0.475[R]$ e podemos verificar que esta proporção é a mesma para a constante que representa o número de monômeros produzidos por transcrição, de forma que $A_{R_2} = 0.475A_R$ que é dada em [15] por $A_R = 11$ monômeros por OC. Em relação às taxas de degradação a proporcionalidade é mantida: $k_{dR_2} = k_{dR}$.

Além disso, na leitura do artigo [15] verificamos ainda que, para simular os equilíbrios dos estados lisogênico e lítico, os autores não levaram em consideração as degradações naturais das proteínas, mas apenas dados parciais *para cada estado em separado*.

Por exemplo, na simulação do estado lisogênico, foram considerados somente valores referentes à produção das proteínas regulatórias e nenhum dado referente à degradação. Neste caso, é evidente que não ocorrerão equilíbrios, pois sem as componentes de degradação as derivadas das funções serão sempre positivas. O que poderá ocorrer será somente uma *diminuição na produção* de proteínas Rep e Cro, mas não um equilíbrio.

Na simulação da indução para o estado lítico, ao contrário do primeiro caso, foi considerada na primeira equação uma taxa alta de degradação para a proteína Rep, provocada por um fator externo (radiação ultravioleta), introduzido em experiências realizadas e que

mostraram ser possível passar do estado lisogênico para o lítico (ver Seção 2.8). Desta forma, há uma grande queda na concentração de dímeros desta proteína Rep, o que permite um aumento na concentração de dímeros da proteína Cro. Com o passar do tempo, poderá ocorrer também uma diminuição na produção de novas proteínas, mas também não ocorrerá um equilíbrio, pois a segunda equação ainda não tem componente negativa e portanto terá derivada sempre positiva.

Mesmo não ocorrendo equilíbrios, percebemos que estas situações acontecem em sistemas distintos. Num primeiro momento, considerando apenas a produção das proteínas regulatórias, simulou-se o estado lisogênico caracterizado pela alta concentração de dímeros da proteína Rep e baixa concentração de dímeros da proteína Cro. E noutro, considerando uma alta taxa de degradação para a proteína Rep, simulou-se o estado lítico, caracterizado pela situação contrária, ou seja, alta concentração de dímeros da proteína Cro e baixa concentração de dímeros da proteína Rep.

Apesar das diferenças entre os modelos (original e deste texto), podemos afirmar que existe coerência entre os mesmos, pois repetindo as simulações do artigo [15] conseguimos reproduzir seus resultados. Só fizemos modificações depois de termos conseguido os mesmos resultados numéricos que os apresentados originalmente.

Tudo isso nos fez refletir sobre a existência de taxas de degradação para ambas as proteínas, e no alto grau de influência que estas taxas exerciam sobre as concentrações.

Suspeitamos então, que poderia haver um valor intermediário, uma taxa de degradação adequada, que pudesse levar um único sistema a exibir dois equilíbrios, e que, a partir deste sistema, a convergência para um ou outro equilíbrio dependesse das condições iniciais das concentrações.

Na leitura de [11], verificamos que há um valor de degradação natural para qualquer proteína presente em uma célula, com exceção da molécula de DNA, que permanece intacta enquanto a célula estiver viva. Esta informação nos permitiu incluir no modelo taxas de degradação para ambas as proteínas e procurar valores para tais taxas de forma que ocorram simultaneamente no modelo os dois equilíbrios. E nos propusemos a estudar a influência destas taxas na dinâmica, por exemplo, investigando a existência ou não de dois equilíbrios estáveis, que representarão os dois estados lisogênico e lítico.

Aqui nos interessa conseguir um sistema que demonstre pelo menos os dois equilíbrios, que representarão os dois estados da bactéria, ou seja, num único retrato de fases deverão aparecer ambos os equilíbrios. Queremos ainda enxergar quais concentrações iniciais das proteínas levarão o modelo para um ou para o outro estado.

4.5 Constantes utilizadas no modelo

Os valores que utilizaremos como constantes foram obtidos, basicamente, de [15]. Este artigo somente não fornecia os valores das taxas de degradação para as proteínas. Esses valores devem ajustar o modelo, de forma que este fique coerente com as características

que apresentamos na Seção 4.4.

Em nossas pesquisas encontramos em [1] valores tidos como taxas de degradação natural para as proteínas Rep e Cro, dadas por

$$\text{Rep} \xrightarrow{k_{dR}} () \text{ onde } k_{dR} = 0.0007s^{-1} ,$$

$$\text{Cro} \xrightarrow{k_{dC}} () \text{ onde } k_{dC} = 0.0025s^{-1} .$$

Como estas informações referem-se às proteínas podemos considerar que esta degradação ocorrerá na mesma proporção tanto para dímeros quanto para monômeros destas proteínas.

Nas simulações, porém, estes valores não nos levaram ao sistema com as características almeçadas. Nos propusemos então a investigar valores mais adequados para estas taxas. Dos resultados obtidos tomamos valores arbitrários, porém com a mesma ordem de grandeza dos apresentados em [1]. Uma discussão mais clara sobre estas simulações e os resultados obtidos serão dados na Seção 5.6.

A Tabela 7 apresenta os valores das constantes que utilizaremos para simulação do modelo.

Observe que estamos considerando a concentração de RNAP, que no modelo foi representada pela letra z , como uma constante.

Existe uma curiosidade sobre a unidade de medida de concentração de moléculas de proteínas do nosso modelo, que se deve ao valor aproximado do volume da bactéria *E.coli*, o que acaba levando a uma certa confusão na leitura de [15].

	P_{RM} e Rep	P_R e Cro	Unidades de medida
$k_{Promotor}$ básico [15]	0.001	0.014	OCs^{-1}
$k_{Promotor}$ estimulado [15]	0.011	—	OCs^{-1}
cte.de proporcionalidade A_{prot} [13, 15]	5.225	6.0	nM OC^{-1}
taxa de degradação	0.0006	0.004	s^{-1}
concentração de RNAP [15]	30	30	nM

Tabela 7: Tabela de parâmetros do modelo

Exibindo explicitamente o valor aproximado do volume da bactéria em *litros* e a unidade de concentração em *Molar = moles por litro*, faremos um cálculo para identificar o valor da concentração na unidade de medida em questão para o caso de apenas 1 molécula de uma proteína presente na bactéria. Representamos a concentração por:

$$\frac{1 \text{ molécula}}{V(\text{volume da bactéria})} = \frac{1 \text{ molécula}}{\pi/2 \cdot 10^{-15} \text{ litros}} = \frac{1 \cdot 10^{15} \cdot 2 \text{ moles}}{6 \cdot 10^{23} \cdot \pi \text{ litros}} ,$$

isto é, aproximadamente,

$$10^{-9} \text{ moles/litro} = 10^{-9} \text{ Molar} = 1 \text{ nanoMolar} = 1 \text{ nM} .$$

Então quando lemos no artigo ou neste trabalho que a concentração de proteínas é de 200 nM, podemos imaginar que existem aproximadamente 200 moléculas de proteínas daquele tipo presentes na bactéria. Por isso, as constantes A_{R_2} e A_{C_2} que são dadas em [15] em número de proteínas por complexos abertos formados, são utilizadas aparentemente nestas mesmas unidades, sem que seja feita uma conversão de número de moléculas para uma unidade de concentração (a não ser a multiplicação por 10^{-9}). Nos cálculos da simulação

do modelo, estes dados são utilizados em molar, então os valores destas taxas, a rigor são, $A_{R_2} = 5.225 \times 10^{-9}$ e $A_{C_2} = 6.0 \times 10^{-9}$ molares por complexos abertos formados. É claro, que se não fosse essa coincidência seria preciso converter número de moléculas em concentração. É importante salientar que a integração numérica foi feita com as concentrações de x e y e demais parâmetros em molares. No entanto, para não carregar a exposição, apresentaremos os resultados em nanomolares (nM).

5 Descrição da dinâmica do modelo

5.1 Análise do modelo

De forma simplificada escreveremos o sistema como

$$\begin{cases} \dot{x} = f(x, y) \\ \dot{y} = g(x, y) \end{cases} .$$

Analisemos agora o comportamento das funções $f(x, y)$ e $g(x, y)$, para dar mais detalhes sobre o campo representado pelo sistema autônomo de equações diferenciais que o modelo representa. Consideraremos que as funções f e g são definidas apenas no quadrante $x \geq 0$ e $y \geq 0$, pois x e y representam quantidades ou concentrações de proteínas. Notemos que f e g são C^∞ e que portanto o campo é dado por um sistema autônomo de classe C^∞ de equações diferenciais ordinárias.

Como o campo é em particular de classe C^1 ficam garantidas a existência e a unicidade das soluções dos problemas de valor inicial.

Nosso primeiro passo será descobrir uma região invariante no primeiro quadrante que, além disso, atrai todas as órbitas do primeiro quadrante. Isso nos permitirá centrar nossa análise apenas nessa região.

5.2 Região invariante

Primeiramente vejamos quais são os comportamentos das funções f e g nos eixos. No eixo x temos $y = 0$, logo a parte negativa de g , correspondente à taxa de degradação, se

anula. Portanto $g > 0$. No eixo y , $x = 0$ e de maneira similar concluimos que $f > 0$. Isso mostra que o primeiro quadrante é invariante pelo fluxo, fato que é coerente com a realidade do fenômeno bioquímico (não há como passar de concentração positiva de proteínas para concentração negativa).

Por outro lado, observando que

$$f(x, y) = [Prob_1[P_{RM}\&RNAP] \cdot k_{PRM1} + Prob_2[P_{RM}\&RNAP] \cdot k_{PRM2}] \cdot A_{R_2} - xk_{dR}$$

sabemos que o primeiro termo é limitado pois é composto por constantes e valores de probabilidades que serão no máximo iguais a 1. Então

$$f(x, y) = [\text{termo limitado}] - xk_{dR} ,$$

portanto existe x_M tal que para todo $x \geq x_M$ teremos $f(x, y) < 0, \forall y \geq 0$.

Analogamente como

$$g(x, y) = Prob[P_R\&RNAP] \cdot k_{PR} \cdot A_{C_2} - yk_{dC}$$

temos que

$$g(x, y) = [\text{termo limitado}] - yk_{dC} ,$$

portanto existe y_M tal que para todo $y \geq y_M$ teremos $g(x, y) < 0, \forall x \geq 0$.

Isso mostra que a região $[0, x_M] \times [0, y_M]$ é invariante pelo fluxo, como mostra a Figura 10.

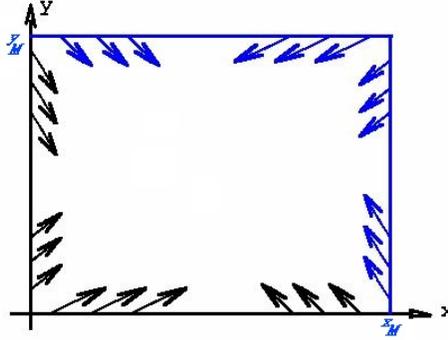


Figura 10: Região invariante para o campo de vetores do modelo

Podemos fixar x_M e y_M com alguma precisão. Na primeira equação, o termo limitado é menor que

$$[k_{PRM1} + k_{PRM2}] \cdot A_{R2} ,$$

logo

$$x_M = \frac{[k_{PRM1} + k_{PRM2}] \cdot A_{R2}}{k_{Rd}}$$

é uma escolha possível. Analogamente

$$y_M = \frac{k_{PR} \cdot A_{C2}}{k_{Cd}} .$$

Além da existência da região invariante, podemos concluir que para qualquer condição inicial (x_0, y_0) existirá um valor t_F tal que

$$(x(t), y(t)) \in [0, x_M] \times [0, y_M]$$

para qualquer $t \geq t_F$.

Como as probabilidades que compõem as funções f e g são proporcionais às concentrações, pode nos parecer contraditório o fato de que para uma alta concentração, ao invés de produzir mais proteínas, a tendência será uma queda nesta concentração.

Mas além da existência de uma taxa de degradação, está embutido no modelo um mecanismo de autoregulação biológico, que prevê um nível de concentração que satura a célula, tornando muito mais provável uma configuração em que não ocorre transcrição. Ou seja, prevê que os sítios operadores serão todos ocupados pela proteína numa situação de alta concentração e isso aumenta consideravelmente o valor da probabilidade de um estado em que não ocorrerá mais a transcrição do gene e, portanto, não será mais produzida, em grande escala, esta proteína.

Tudo isso vale tanto para a proteína Rep, quanto para a Cro pois, no caso do estado lítico, ocorre uma queda acentuada na concentração de dímeros da proteína Rep, e um aumento na concentração de dímeros da proteína Cro. Com o passar do tempo, se repetirá a situação anterior de alta concentração de dímeros, só que agora da proteína Cro.

Portanto, se encontrarmos o retrato de fases dentro desta região invariante saberemos basicamente o comportamento do sistema no quadrante como um todo.

Para os parâmetros apresentados na Tabela 7 a discussão da região invariante pode ser

estimada por

$$\begin{aligned}x_M &= \frac{[k_{PRM1} + k_{PRM2}] \cdot A_{R2}}{k_{Rd}} \\ &= \frac{[0.011 + 0.001] \cdot 4.4 \cdot 10^{-9}}{0.0006} \\ &= 104.5 \times 10^{-9} \text{ Molares ,}\end{aligned}$$

e

$$\begin{aligned}y_M &= \frac{k_{PR} \cdot A_{C2}}{k_{Cd}} \\ &= \frac{0.014 \cdot 6.0 \cdot 10^{-9}}{0.004} \\ &= 21 \times 10^{-9} \text{ Molares .}\end{aligned}$$

Podemos concluir por estes cálculos que, no quadrante positivo, não existirão singularidades fora da região $[0 \text{ nM}; 104.5 \text{ nM}] \times [0 \text{ nM}; 21 \text{ nM}]$. Então, se existirem singularidades no quadrante $x \geq 0$ e $y \geq 0$ certamente estarão dentro desta região.

5.3 Encontrando singularidades

As singularidades e demais propriedades da equação diferencial foram encontrados com o auxílio do software Maple V.

Foi utilizado aqui o comando *solve* com as equações $f(x, y) = 0$ e $g(x, y) = 0$ para encontrar suas raízes e foram obtidos valores reais positivos, reais negativos e complexos. Mas os resultados que nos interessam são apenas os reais positivos pois representam *singularidades*

reais positivas:

$$P_1 = (4.93243475 \text{ nM}, 16.29916562 \text{ nM}) ,$$

$$P_2 = (11.88831310 \text{ nM}, 12.57120402 \text{ nM}) ,$$

$$P_3 = (44.95477144 \text{ nM}, 3.02513962 \text{ nM}) .$$

Estas três raízes são aproximações de três singularidades, e coerentemente estão dentro da região invariante. A existência de três singularidades dentro da região invariante é confirmada visualmente quando traçamos as curvas de nível zero das funções f e g .

As curvas de nível das funções f e g revelam seu comportamento (em geral) decrescente, como mostram as Figuras 11–A e 11–B. Tomando apenas as curvas de nível zero, confirmamos que ocorrerem três cruzamentos, portanto três locais em que o campo se anula (Figura 11–C).

Pelo estudo dos casos de competições entre espécies, como é feito em [7], sabemos que as curvas de nível zero de f e g têm propriedades importantes no que diz respeito ao campo de vetores. Sobre a curva $f^{-1}(0)$ a componente \dot{x} se anula portanto somente a componente \dot{y} aparecerá, o vetor terá direção vertical e apontará para cima ou para baixo dependendo do sinal de \dot{y} . O mesmo ocorre sobre a curva $g^{-1}(0)$ sobre a qual somente a componente horizontal existirá e seu sentido (direita ou esquerda) dependerá do sinal de \dot{x} . A Figura 11–D ilustra esses fatos.

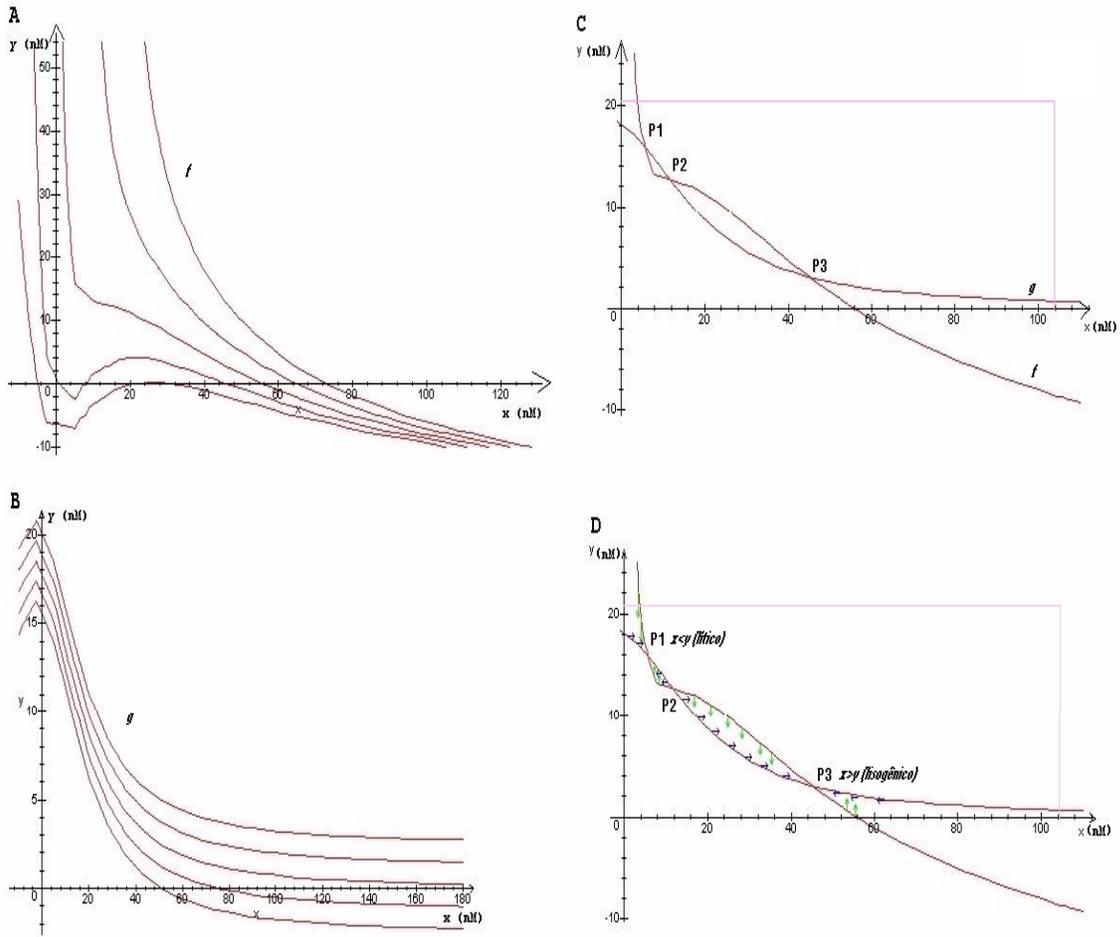


Figura 11: A - Gráfico que representa curvas de nível da função f ; B - Gráfico que representa as curvas de nível da função g ; C - Gráfico que mostra as curvas de nível 0 das funções f e g , dentro da região invariante; D - Gráfico com curvas de nível 0 das funções f e g com as direções do campo sobre elas.

5.4 Classificação das singularidades

Que tipo de singularidades são estas? Como se comporta o fluxo em suas vizinhanças?

Dizemos que uma singularidade p de um campo X é *hiperbólica* se $DX(p)$ é um campo linear hiperbólico, isto é, não tem autovalores com parte real nula. O Teorema de Hartman-Grobman (ver, por exemplo, [3]) para campos de vetores garante que se p é ponto fixo hiperbólico de X então, X em p é localmente topologicamente conjugado a $DX(p)$ em 0.

Analisemos primeiro $P_1 = (4.93243475, 16.29916562)$. Verifiquemos primeiramente se é *hiperbólico*. A matriz Jacobiana $DX(P_1)$, calculada com o auxílio do Maple, é

$$\begin{pmatrix} -0.0002349103828 & -0.0001013939616 \\ -0.002154227390 & -0.004456792506 \end{pmatrix},$$

cujos autovalores são

$$\lambda_1 = -0.0001837927461, \lambda_2 = -0.004507910143.$$

Seus autovetores são:

$$v_1 = (0.09987315877, -0.05035092585), v_2 = (0.02347825579, 0.9894334905).$$

Uma aproximação do retrato de fases local, ou seja, nas proximidades desta singularidade pôde ser exibido com os recursos dos softwares já citados (ver Figura 12).

Portanto P_1 é um nó atrator hiperbólico e o campo X em P_1 é localmente conjugado com o campo linear $DX(P_1)$ em 0 (Teo Hartmann-Grobman). Podemos notar ainda que

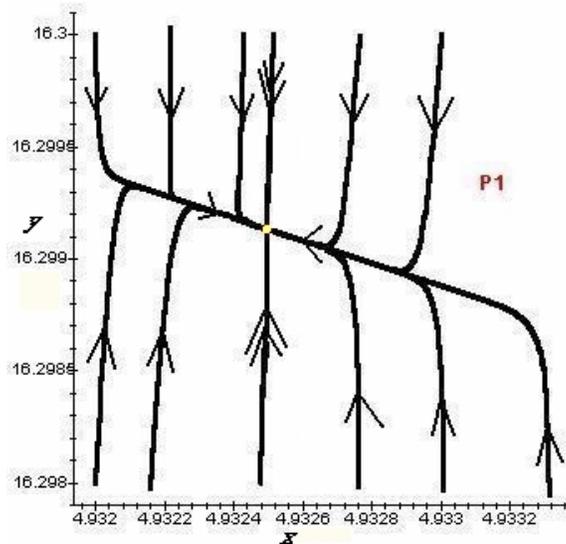


Figura 12: Retrato de fases local - vizinhança da singularidade P_1

uma das direções atrai de maneira muito mais forte que a outra direção: isso se dá pois existe uma diferença na ordem de grandeza dos autovalores de aproximadamente cem vezes. Como o valor absoluto do autovalor está associado à velocidade de convergência, ocorrerá uma aproximação com relação à singularidade de forma tangente à direção mais fraca.

Para sermos coerentes com os dados expostos até aqui, podemos afirmar que este equilíbrio refere-se ao *estado lítico* ou virulento, pois a concentração de dímeros da proteína (y , Cro) que estão associadas a este estado é muito maior que a concentração da outra proteína (x , Rep).

Vejamos agora o que ocorre com $P_2 = (11.88831310, 12.57120402)$. Temos a matriz

$DX(P_2)$ igual a

$$\begin{pmatrix} -0.0000202581096 & -0.0002899596442 \\ -0.002350907575 & -0.004335179747 \end{pmatrix},$$

cujos autovalores são

$$\lambda_1 = 0.0001323255536, \lambda_2 = -0.004487763410$$

e cujos autovetores são

$$v_1 = (0.8849517749, -0.4656826776), v_2 = (0.0709198185, 1.092685383).$$

Trata-se de uma sela hiperbólica, onde uma das direções atrai e a outra repele. Pela mesma análise feita anteriormente da força de atração e/ou repulsão para o nó hiperbólico, através da ordem de grandeza dos autovalores percebemos que a direção de atração é muito mais forte que a direção de repulsão, o que explica a grande deformação que podemos ver no retrato de fases local (Figura 13).

Este equilíbrio não tem, em princípio, significado no fenômeno bioquímico, embora logo adiante fique claro o papel desta sela na divisão das bacias de atração de cada equilíbrio estável.

Finalmente, para $P_3 = (44.95477144, 3.02513962)$ temos a matriz

$$\begin{pmatrix} -0.000457228099 & -0.00148420918 \\ -0.000436329853 & -0.004037625879 \end{pmatrix},$$

cujos autovalores são

$$\lambda_1 = -0.0002846695302, \lambda_2 = -0.004210184448$$

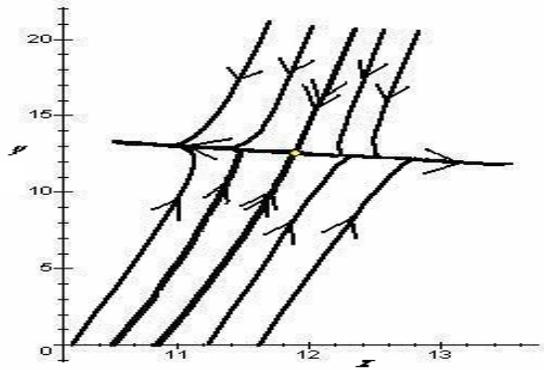


Figura 13: Retrato de fases local - vizinhança da singularidade P_2

e cujos autovetores são

$$v_1 = (0.9933092149, -0.1154850798) , v_2 = (0.3806396443, 0.9624815576) .$$

Este equilíbrio (Figura 14) é associado ao *estado lisogênico* pelos mesmos argumentos que foram dados anteriormente para P_1 .

Dentro da região invariante existem estas três singularidades, todas hiperbólicas e portanto simples (pontos são isolados).

5.5 O retrato de fases

Utilizando ainda os recursos dos softwares podemos então esboçar uma aproximação do retrato de fases desta região, como mostra a Figura 15. Note que a variedade estável da sela P_2 é a fronteira das bacias de atração dos pontos atratores hiperbólicos P_1 e P_3 . Para encontrá-la, verificamos primeiro que ela deveria intersectar as linhas $y = 0$ e $y = y_M$

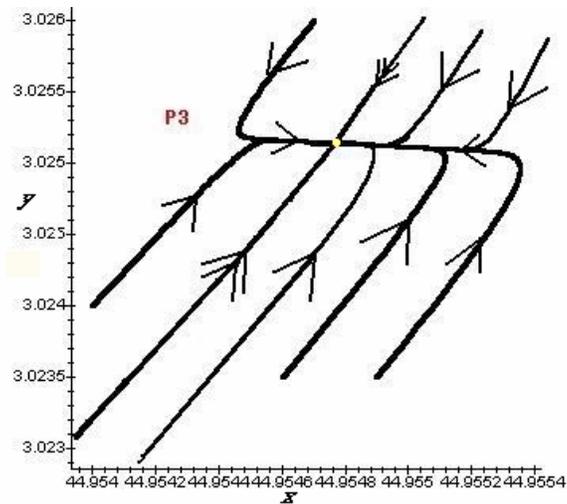


Figura 14: Retrato de fases local - vizinhança da singularidade P_3 .

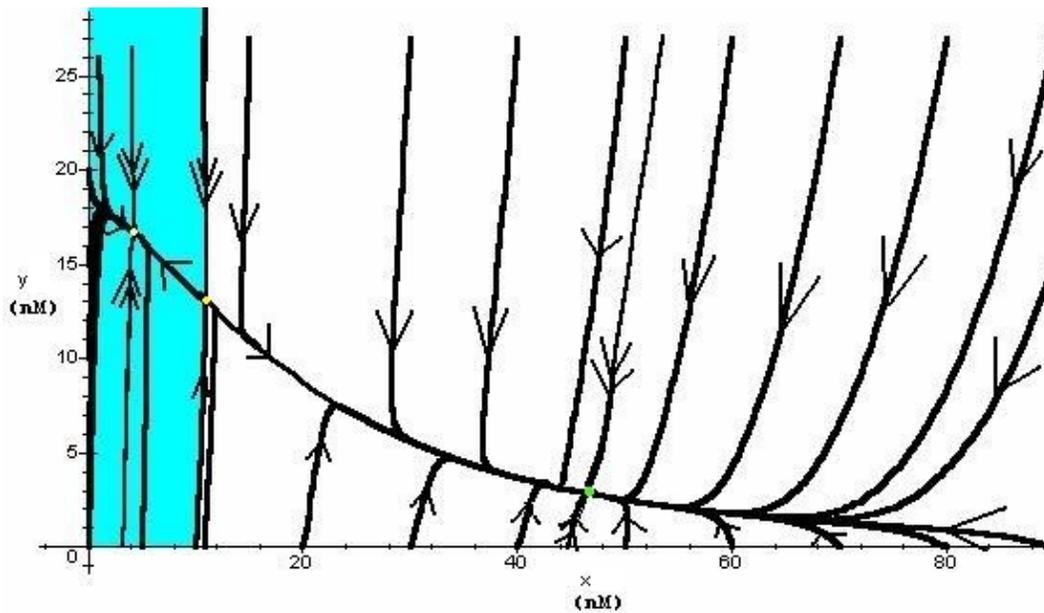


Figura 15: Retrato de fases dentro da região invariante para o campo de vetores do modelo. A região sombreada é a bacia de atração de P_1 , delimitada pelo eixo y e pela variedade estável da sela P_2 .

que ajudam a definir a região invariante. Variando x_0 , mantendo $y_0 = 0$ ou $y_0 = y_M$ e usando aproximação por dicotomia, conseguimos reproduzir duas trajetórias com proximidade bastante confiável das separatrizes da variedade estável da sela. Não foi necessária uma integração do campo com sinal invertido, pois nossas condições iniciais já se encontravam nas fronteiras (tanto superior como inferior) da região invariante, que é exatamente a área que nos interessa.

Aparentemente também se depreende, da Figura 15, que existe uma linha invariante que atrai todas as órbitas e que contém as três singularidades.

Este retrato de fases nos parece um resultado coerente com os dados encontrados nas referências, onde existem muito mais bactérias no estado lisogênico do que no lítico, já que cada estado está associado às concentrações das proteínas regulatórias. De acordo com o retrato de fases, podemos verificar que o equilíbrio P_3 tem uma bacia de atração muito maior que a do outro equilíbrio P_1 . Portanto este modelo determinístico, utilizando os valores de parâmetros citados, parece-nos coerente e viável.

5.6 Comentários sobre os parâmetros

Também determinamos, através de simulação em computador, a região de pares de taxas de degradação das proteínas para os quais o retrato de fases fica semelhante ao apresentado na Figura 15. Através de um estudo sobre as curvas de nível zero e do cálculo direto das raízes das funções, verificamos que existirão 2 situações distintas no modelo:

1. A situação já descrita, onde aparecerão 3 singularidades; e
2. uma situação onde aparecerá apenas uma singularidade.

Pelo menos numericamente, parece que a inexistência de singularidades só ocorre quando uma das taxas de degradação é nula. A Figura 16 mostra o resultado. A grande maioria de pares de valores refere-se ao aparecimento de uma única singularidade, embora ali não estejam especificadas regiões em separado para uma única singularidade referente ao estado lítico, ou referente ao estado lisogênico. E aparece ainda uma região onde os pares de taxas de degradação proporcionam o surgimento de três singularidades, como é o caso do nosso modelo.

Conforme vão aumentando (ambas) as taxas de degradação a faixa que corresponde ao surgimento de três singularidades vai se estreitando até sumir. Observamos ainda que, no espaço de fases, a região invariante pode ser tomada cada vez menor e próxima da origem. Na extremidade, onde a faixa escura do gráfico da Figura 16 termina, as simulações nos mostram que a região invariante no primeiro quadrante é muito pequena e as singularidades estão muito próximas. Para esses parâmetros, mesmo que ainda haja uma pequena região invariante e singularidades dentro dela, o número de proteínas envolvidas é muito pequeno e é difícil tirar uma conclusão sobre a situação biológica do sistema.

Uma observação atenta ao gráfico da Figura 16 nos permite localizar o par de taxas que utilizamos nas nossas simulações e ainda verificar que as taxas de degradação citadas

na Seção 4.5 estão na região referente a uma única singularidade (estado lítico), mas bem próxima da região que nos interessa, ou seja, daquela em que aparecem as três singularidades. Isso nos faz refletir sobre as diversas, aproximações feitas ao longo deste trabalho. Por exemplo, no cálculo do volume da bactéria, cuja aproximação afetará diretamente as constantes A_{R_2} e A_{C_2} e ainda nos próprios valores dos parâmetros que devem ter desvios. Não há nenhuma razão para se descartar a hipótese de que no sistema biológico real os parâmetros sejam tais a permitir a coexistência dos dois equilíbrios dinâmicos.

A transição no espaço de parâmetros entre as duas situações qualitativamente distintas se dá através da tangência entre as curvas de nível 0, o que corresponde também à colisão entre um nó hiperbólico e uma sela hiperbólica. Esse tipo de bifurcação é conhecido como *bifurcação de sela-nó* ([7]). Podemos imaginar que, biologicamente isso corresponderia a introduzir o raio UV no sistema, aumentando assim a taxa de degradação da proteína Rep e portanto levar o sistema do equilíbrio lisogênico ao lítico.

Tomando três valores distintos de taxas de degradação para a proteína Rep, mantendo a outra fixa em $k_{dC} = 0.004$ temos caracterizadas duas situações de transferência. Num primeiro momento, com uma taxa de degradação baixa, $k_{dR} = 0.0005$ encontramos somente um equilíbrio, que corresponde ao estado lisogênico, pois a concentração de dímeros da proteína do tipo Rep é muito maior que a concentração de dímeros da outra proteína, a Cro. Em seguida, considerando a taxa de degradação $k_{dR} = 0.0006$, a mesma utilizada na

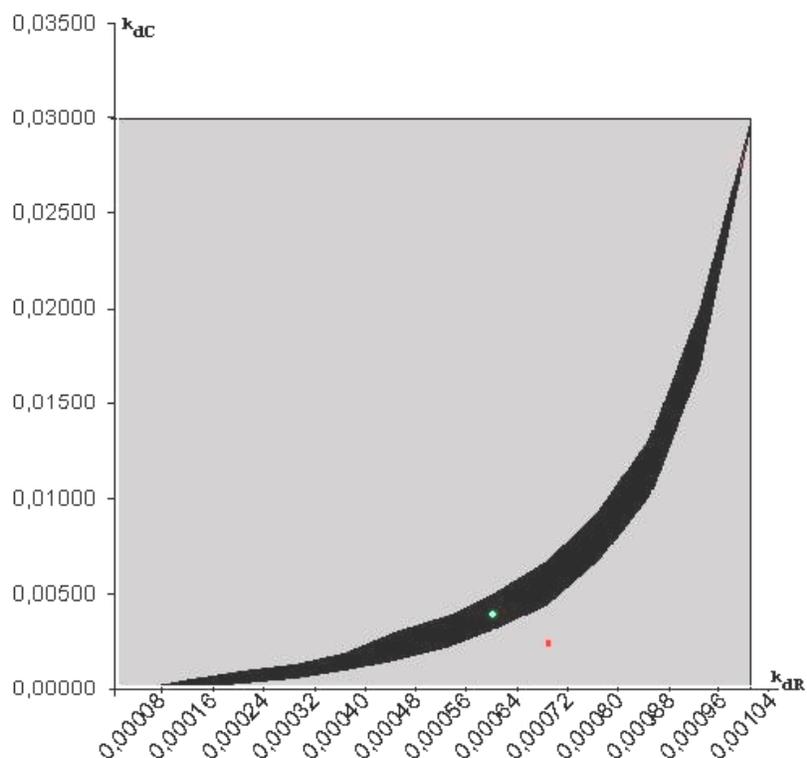


Figura 16: Gráfico da região de pares de taxas de degradação. Área cinza - apenas uma singularidade (à esquerda, lisogênico, e à direita, lítico). Área escura - três singularidades. Ponto no interior da área escura - par de taxas utilizado no modelo.

simulação do modelo, encontramos os três equilíbrios que já foram descritos anteriormente.

E por último, considerando uma taxa de degradação mais elevada $k_{dr} = 0.00075$, aparece novamente somente um equilíbrio, com a diferença de que agora este corresponde ao estado lítico, já que a concentração de dímeros da proteína Cro que é a maior agora.

Podemos concluir então que existirá um valor para k_{dr} entre 0.0005 e 0.0006 para o qual aparecerão dois equilíbrios. Se esta taxa vai aumentando continuamente, para algum

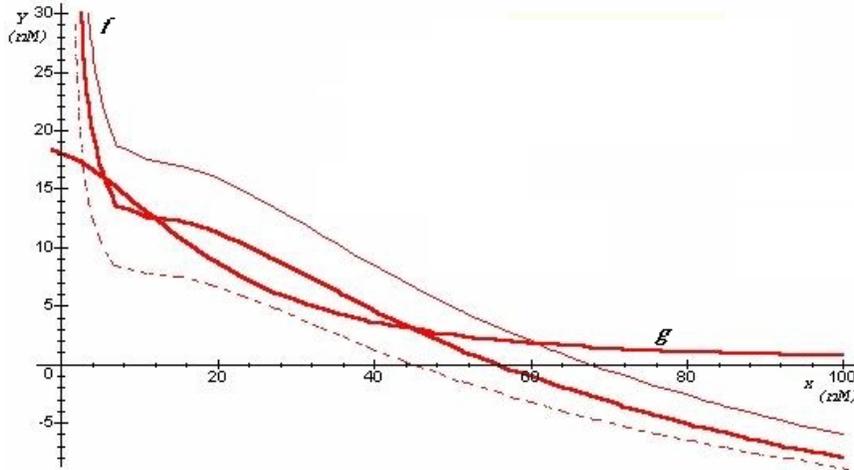


Figura 17: Curvas de nível 0 de funções f e g com diferentes taxas de degradação. Fixando-se a taxa de degradação $k_{dC} = 0.004$ e variando a taxa de degradação k_{dR} percebemos que quanto maior o valor desta taxa, mais rapidamente decresce a função f . Ao considerarmos três taxas de degradação k_{dR} distintas para traçar as curvas de nível 0, constatamos a ocorrência de três situações: $k_{dR} = 0.0006$, (taxa do modelo - linhas mais espessas) três singularidades já descritas. Para $k_{dR} = 0.0005$ (mais baixa que a do modelo - linha contínua fina) uma singularidade correspondente ao estado lisogênico. Para $k_{dR} = 0.00075$ (mais alta que a do modelo - linha pontilhada) uma singularidade correspondente ao estado lítico.

valor entre os dois já citados, as curvas de nível zero das funções, que se cruzavam uma única vez, passam a se cruzar duas vezes, e logo após, cruzam-se três vezes, como mostra a Figura 17 e a Figura 18, que mostra também, esquematicamente, a configuração dinâmica dos equilíbrios.

Da mesma forma, existe um outro valor entre 0.0006 e 0.00075 em que novamente existirão somente dois equilíbrios. Nessas duas situações onde há exatamente dois equilíbrios, um deles é uma sela-nó.

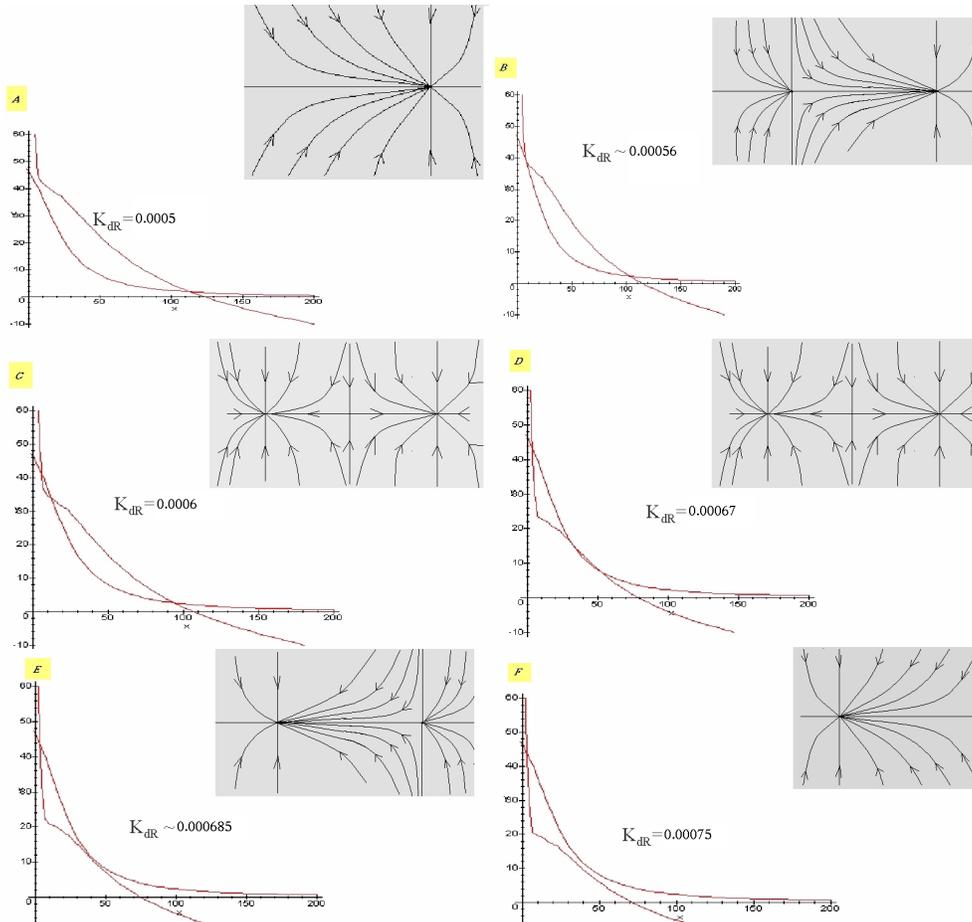


Figura 18: Gráficos com curvas de nível zero das funções f e g . Taxa de degradação k_{dC} fixa em $0.004s^{-1}$ A - Gráfico com $k_{dR} = 0.0005s^{-1}$, um equilíbrio (lisogênico); B - Gráfico com $k_{dR} \sim 0.00056s^{-1}$, dois equilíbrios com ocorrência de sela-nó; C - Gráfico com $k_{dR} = 0.0006s^{-1}$, três equilíbrios (semelhante ao modelo apresentado); D - Gráfico com $k_{dR} = 0.00067s^{-1}$, três equilíbrios (semelhante ao modelo apresentado, porém com dimensões diferentes para as bacias de atração); E - Gráfico com $k_{dR} \sim 0.000685s^{-1}$, dois equilíbrios com ocorrência de sela-nó; F - Gráfico com $k_{dR} = 0.00075s^{-1}$, um equilíbrio (lítico).

6 Considerações finais

Através de considerações de caráter físico-químicas, como foi feito em [15], obtivemos um sistema de equações diferenciais em duas variáveis que modela a dinâmica das concentrações das proteínas Rep e Cro numa bactéria invadida pelo vírus fago- λ . Nossa modelagem incorporou certas modificações ao modelo originalmente proposto, ao usar taxas de degradação para ambas as proteínas, de forma que os dois estados, lisogênico e lítico, aparecessem como pontos de equilíbrio de *um mesmo modelo*. Já nas simulações feitas no artigo original ([15]) não ocorreram realmente equilíbrios, mas apenas uma queda considerável na taxa de produção das proteínas. Isso resultou numa pseudo-estabilização na concentração de proteínas - o aumento é tão pequeno que não é percebido num espaço curto de tempo, dando-nos uma falsa idéia de estabilidade na concentração. Mas a inclusão de taxas de degradação para ambas as proteínas proporcionou ao modelo uma melhor aproximação com a realidade e a possibilidade de um estudo mais preciso.

O equilíbrio referente ao estado lisogênico que constatamos pelo modelo é também um equilíbrio biológico, pois a bactéria pode permanecer neste estado por um longo tempo. Já o equilíbrio referente ao estado lítico é estável do ponto de vista matemático, mas biologicamente sabemos que, este estado é extremamente nocivo à bactéria. Logo que entra neste estado, desencadeia um processo de fabricação de novos vírus, o λ -DNA se solta do DNA da bactéria e acaba por matá-la, destruindo assim o sistema como um todo.

Além disso, durante este processo várias moléculas de λ -DNA são fabricadas, fazendo com que o modelo perca a validade. Já a bacia de atração do equilíbrio lisogênico é um dado quantitativo válido, pois está na região onde as hipóteses utilizadas têm validade.

Algumas características biológicas reais não foram consideradas neste modelo, como por exemplo, o fato de ocorrerem flutuações nas concentrações das proteínas, ou ainda, a presença momentânea de duas moléculas de DNA por ocasião da reprodução da bactéria. Essas informações podem mudar a dinâmica do sistema, e desta forma, para verificarmos como será seu comportamento, precisamos incorporar estas informações ao modelo e refazer toda a simulação.

Há ainda, uma observação importante sobre o modelo que gostaríamos de dividir com os leitores. Observamos que a configuração 16, em que duas RNAP estão ligadas ao mesmo tempo nos sítios promotores de ambos os genes, é utilizada em ambas as equações, isso porque a expressão da probabilidade conta as configurações em que a RNAP está presente num sítio promotor - cada equação considerando o sítio promotor do gene correspondente à proteína à qual se refere. Ao multiplicarmos a probabilidade que engloba esta configuração pela constante de proporcionalidade que dá o número de proteínas produzidas por transcrição, temos a impressão de que ocorrerá transcrição para os dois promotores ao mesmo tempo, de fato é como se admitíssemos que isto ocorre. Porém há informações em [13] e [15] que indicam o contrário. Nos reservaremos aqui ao direito de continuar utilizando

o modelo formulado, apesar de ter levantado uma questão que julgamos importante. Em trabalhos posteriores achamos importante que essa informação seja averiguada.

Acreditamos também termos dado uma contribuição na facilidade de acesso ao modelo, expondo lado a lado os aspectos qualitativos, como em [13] e os quantitativos dados em [15], e ainda oferecendo uma explicação para as probabilidades das configurações (Seção 4.1 e Equação 15) que não encontramos em nenhum dos textos pesquisados. A exposição sobre Termodinâmica Estatística e Equilíbrio Químico serve como primeira apresentação aos não iniciados nessas disciplinas, que podem assim usar ou questionar o modelo com mais propriedade.

As adaptações feitas no modelo tiveram como objetivo dar um enfoque mais matemático ao assunto, permitindo-nos uma breve comparação com um modelo populacional de competição entre espécies, além de utilizar dados mais realistas para a questão - como as taxas de degradação natural para as duas proteínas, algo que realmente existe como confirmado em bibliografia adequada.

Algumas questões ainda podem ser levantadas como: este tipo de modelo é geral em expressão gênica? Se introduzirmos mais proteínas no problema, ou considerarmos um outro problema em que mais proteínas estão envolvidas poderemos atingir uma situação de caos? Por exemplo, considerando um sistema em \mathbb{R}^3 (mas poderia ser em dimensão n)

do tipo

$$\begin{cases} \dot{x}_1 = \frac{\sum_i b^{i1} x_1^{j_i} x_2^{k_i} x_3^{m_i}}{\sum_i a_i x_1^{j_i} x_2^{k_i} x_3^{m_i}} - k_1 x_1 \\ \dot{x}_2 = \frac{\sum_i b^{i2} x_1^{j_i} x_2^{k_i} x_3^{m_i}}{\sum_i a_i x_1^{j_i} x_2^{k_i} x_3^{m_i}} - k_2 x_2 \\ \dot{x}_3 = \frac{\sum_i b^{i3} x_1^{j_i} x_2^{k_i} x_3^{m_i}}{\sum_i a_i x_1^{j_i} x_2^{k_i} x_3^{m_i}} - k_3 x_3 \end{cases} .$$

com $b^{in} \geq 0$, $a_i \geq 0$ e $k_i \geq 0$, que é formulado nos mesmos moldes que o apresentado neste trabalho, porém um pouco mais geral, composto por três variáveis e constantes positivas arbitrárias. Que tipos de comportamento podem ou não podem ter? Sabemos o quão complexo pode se tornar um sistema qualquer em \mathbb{R}^3 , e a questão que colocamos é justamente se podem surgir novos elementos além das singularidades, como no caso particular que apresentamos.

Mesmo em dimensão 2, pode-se buscar saber que tipos de comportamento esse tipo de modelo proporciona. Por exemplo, pode haver mais equilíbrios estáveis? Como eles se distribuem? É um fato geral a presença de uma linha invariante que contém todas as singularidades e que atrai todas as órbitas?

Para uma investigação bioquímica poderiam-se fazer experimentos baseados nas Figuras 15 e 16. No caso da primeira, pode ser interessante confirmar o posicionamento da bacia de atração do estado lisogênico e as trajetórias até esse equilíbrio. No caso da segunda, estudar como as taxas de degradação afetam a existência dos equilíbrios e como mudam as bacias de atração.

Referências

- [1] Arkin, Adam; Ross, John; MacAdams, Harley H. “Stochastic kinetic analysis of developmental pathway bifurcation in phage λ -infected *Escherichia coli* cells.” *Genetics*, Baltimore, v.149, n. 4, p. 1633-1648, 1998.
- [2] Callen, Herbert B. *Thermodynamics and an introduction to thermostatistics. 2.ed.* New York: John Wiley, 1985. 493p.
- [3] Fichmann, Luiz; Sallum, Élvia M. *Sistemas dinâmicos, noções básicas.* São Paulo: Publicações de IME-USP, 2004. 184p.
- [4] Gasiorowicz, Stephen. *Física Quântica.* Ed. Guanabara, Rio de Janeiro, 1979. 480p.
- [5] Gibson, Michael A.; Bruck, Jehoshua. “A probabilistic model of a prokaryotic gene and its regulation.” In: Bolouri, Hamid; Bower, James M. eds. *Computational modeling of genetic and biochemical networks.* Cambridge, MIT Press, 2001. p. 49-72.
- [6] Hill, Terrell L. *Lectures on matter and equilibrium.* New York: W. A. Benjamin, 1966. 306p.
- [7] Hirsch, Morris W.; Smale, Stephen. *Differential equations, dynamical systems, and linear algebra.* New York: Academic Press, 1974. 358p. (Pure and Applied Mathematics; Series of Monographs and Textbooks, 60)

- [8] Lima, Elon Lages. *Curso de análise vol. 2 (6ª edição)*. IMPA, Rio, 2000. 487p.
- [9] Lopes, Sônia G. B. C. *Bio: introdução à biologia e origem da vida, citologia, reprodução e embriologia, histologia. 2.ed.* São Paulo: Saraiva, 2003. 466p.
- [10] MacAdams, Harley H.; Shapiro, Lucy. "Circuit simulation of genetic networks." *Science*, v. **269**, p. 650-655, 1995.
- [11] Mahan, Bruce H. *Química - um curso universitário. 2.ed.* São Paulo: Edgard Blucher, 1972. 654p.
- [12] Mortimer, Robert G. "Physical chemistry." Redwood City: Benjamin/Cummings Pub., c1993. 1014p. (The Benjamin/Cummings Series in the Life Sciences and Chemistry)
- [13] Ptashne, Mark. *A genetic switch: phage λ and higher organisms. 2.ed.* Cambridge: Cell Press: Blackwell Scientific Publications, 1992. 192p.
- [14] Reif, Frederick. *Fundamentals of statistical and thermal physics*. New York: McGraw Hill, 1965. 651p.
- [15] Shea, Madeline A.; Ackers, Gary K. "The control Or system of bacteriophage λ - a physical-chemical model for gene regulation." *Journal of Molecular Biology*, v. **118**, p. 211-230, 1985.