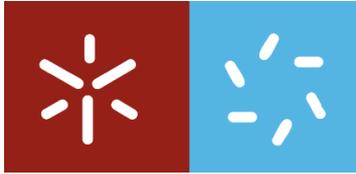


**Universidade do Minho**  
Escola de Ciências

Nuno Elmano Moutinho Lima

**Comparação de métodos computacionais  
para estimar a energia livre de Gibbs  
em reacções bioquímicas**



Universidade do Minho  
Escola de Ciências

Nuno Elmano Moutinho Lima

Comparação de métodos computacionais para  
estimar a energia livre de Gibbs em reações  
bioquímicas

Dissertação de Mestrado

Mestrado em Biofísica e Bionanossistemas

Trabalho efetuado sob a orientação de

**Professor Doutor Luís Silvano Alves Marques**

Outubro de 2018

## DECLARAÇÃO

Nome: Nuno Elmano Moutinho Lima

Endereço eletrónico: nunoelima@gmail.com Telefone: 927 710 842

Cartão do Cidadão: 13336029

Título da dissertação: Comparação de métodos computacionais para estimar a energia livre de Gibbs em reações bioquímicas

Orientador:

Professor Doutor Luís Silvino Alves Marques

Ano de conclusão: 2018

Mestrado em Biofísica e Bionanossistemas

É AUTORIZADA A REPRODUÇÃO INTEGRAL DESTA DISSERTAÇÃO APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE.

Universidade do Minho, \_\_\_\_/\_\_\_\_/\_\_\_\_

Assinatura:

Para a Eva.

*“Acorda menina linda,  
vem oferecer  
o teu sorriso ao Dia  
que acabou de nascer.”*

Para a Carla.

*“Os romances maus contam histórias,  
os bons romances mostram-nos a nós mesmos.”*

## AGRADECIMENTOS

---

Um “muito obrigado!” ...

...ao professor Luís Marques, por ter aceitado um orientando ausente, que foi acumulando outras preocupações e que talvez nunca tenha mostrado garantias de levar isto a bom porto.

...aos meus pais e à minha irmã, por tanto terem insistido para que não deixasse esta etapa por concluir e... bem, por serem a melhor família que se pode desejar.

...aos *Patifes*, porque não importa o tempo nem a distância. É sempre como se tivesse sido ontem.

...à ARCUM e à Tuna Universitária do Minho, por me terem permitido construir muito mais do que um percurso universitário.

...às pessoas a quem dedico esta dissertação, a Eva e a Carla, por criarem em mim a necessidade de ser melhor.

## RESUMO

---

A energia livre de Gibbs, por indicar o saldo energético de uma reação química ou bioquímica, fornece indicações acerca da direção e da reversibilidade que essa reação e, em última instância, uma via metabólica completa, podem tomar. Estas informações podem ser de uma utilidade crucial na concepção de processos bioquímicos da maior relevância quer em investigação científica, quer ao nível de aplicações inovadoras nas mais variadas áreas da biotecnologia.

No entanto, a disponibilidade de dados termodinâmicos concretos e, em especial, sobre a variação da energia livre de Gibbs, em reações do metabolismo celular é escassa e, frequentemente, incompleta. Ademais, os constrangimentos práticos da sua obtenção pela via experimental impedem a avaliação termodinâmica de grande número de reacções.

Os métodos estimativos são por isso uma opção atrativa, tendo já sido publicados por vários autores diversos trabalhos que visam o desenvolvimento de formas alternativas de avaliar a variação da energia livre de Gibbs de um elevado número de reacções num curto espaço de tempo. Os esforços mais recentes e promissores neste sentido procuram combinar, de uma forma completa e consistente, os dados obtidos pela via experimental com diferentes abordagens teóricas sobre a previsibilidade da termodinâmica ao nível de reacções bioquímicas.

Nesta dissertação, colocam-se em destaque dois métodos em particular. O primeiro, *Component Contribution*, utiliza dados provenientes de trabalhos experimentais, aliados a outras informações teóricas e a um princípio assente na decomposição das espécies químicas envolvidas em vários grupos funcionais. O segundo, por sua vez, baseia-se na ideia de que os dados experimentais disponíveis na literatura podem ser aplicados segundo uma lógica de semelhanças estruturais e transformacionais para inferir dados acerca de reacções para as quais os parâmetros termodinâmicos são ainda desconhecidos. Na impossibilidade de aplicar diretamente este segundo método nos seus moldes originais, neste trabalho optou-se por desenvolver de raiz um programa baseado nos mesmos princípios teóricos. São ainda tecidas considerações sobre o âmbito de aplicabilidade de cada um dos métodos abordados e avaliada a capacidade de um poder complementar o outro, numa ferramenta mais completa e mais robusta.



## ABSTRACT

---

For its ability to denote the energetic balance of a chemical or biochemical reaction, Gibbs free energy provides information about the direction and reversibility that it, or even a complete metabolic pathway, can take. Such information may be of great value for designing biochemical processes of huge relevance for both scientific research and innovative applications in the most diverse fields of biotechnology.

However, availability of specific data about the thermodynamics – and especially about the Gibbs free energy – of metabolic reactions is limited and often incomplete. Moreover, practical constraints involved in their experimental measurement prevent ascertaining thermodynamical parameters for a large number of reactions.

Estimation methods pose therefore an attractive option, and several works have been published envisioning the development of alternative ways for evaluating the change in Gibbs free energy for large sets of reactions in a short time. The most recent and promising efforts look for consistently combining experimental data and different theoretical approaches on predicting thermodynamics in biochemical reactions.

This thesis focuses on two methods. The first one, *Component Contribution*, uses data from experimental works together with other theoretical information and the idea that chemical species taking part in a reaction can be decomposed into several functional groups. The second one relies on the principle that the available experimental data can be used for searching structural and transformational similarities for inferring data for reactions with unknown thermodynamic parameters. Although the second method could not be implemented in its original formulation, another method was developed from scratch during this work, using the same theoretical principles employed in the original one. The practical applicability of each method was evaluated, as well as the possibility of their implementation in a complementary manner.



# ÍNDICE

<b>Agradecimentos</b> .....	iv
<b>Resumo</b> .....	v
<b>Abstract</b> .....	vii
<b>Índice de Figuras</b> .....	xi
<b>Índice de Tabelas</b> .....	xiii
<b>1. Introdução</b> .....	1
<b>1.1. A termodinâmica das reações bioquímicas e a energia livre de Gibbs<sup>[21]</sup></b> .....	5
1.1.1. A primeira lei da termodinâmica.....	5
1.1.2. Segunda lei da termodinâmica.....	6
1.1.3. Energia livre de Gibbs e processos espontâneos.....	8
1.1.4. Variação da energia livre de Gibbs no estado de equilíbrio.....	10
1.1.5. Convenções sobre condições-padrão.....	12
<b>1.2. Estimativa da energia livre de Gibbs por métodos computacionais</b> .....	13
1.2.1. Contribuições de grupos funcionais – ‘ <i>Group Contribution</i> ’.....	15
1.2.2. Contribuições de Reagentes – ‘ <i>Reactant Contribution</i> ’.....	16
1.2.3. Contribuições de componentes - ‘ <i>Component Contribution</i> ’.....	18
1.2.4. IGERS – ‘ <i>Inferring Gibbs Energy changes from Reaction Similarities</i> ’.....	22
1.2.5. Outros métodos computacionais.....	25
<b>1.3. Escolha dos métodos a analisar</b> .....	29
<b>1.4. Disponibilidade de IGERS e métodos alternativos</b> .....	31
<b>1.5. Objetivos</b> .....	34
<b>2. Descrição das metodologias adotadas e das vias metabólicas estudadas</b> .....	35
<b>2.1. A base de dados NIST-TECRDB</b> .....	37
2.1.1. Base de dados TECR em <i>Component Contribution</i> – linearização e conjugação com informações adicionais.....	38
2.1.2. Utilização de TECR em <i>Reaction Fingerprinter</i> .....	42
<b>2.2. <i>Reaction Fingerprinter</i> – JChem®, ‘<i>Reaction Fingerprints</i>’ e ‘<i>Similarity Search</i>’</b> .....	43
2.2.1. Desenvolvimento de <i>script</i> para gerar ficheiros <i>RXN</i> .....	44
2.2.2. <i>Similarity search</i> e <i>reaction fingerprints</i> .....	47
<b>2.3. Seleção da amostra de 200 reações metabólicas</b> .....	51
<b>2.4. Seleção de um grupo de reações com <math>\Delta G_R^0</math> conhecida mas ausentes de TECR</b> .....	53

<b>2.5. Descrição das vias metabólicas para análise por <i>Component Contribution</i> e <i>Reaction Fingerprinter</i></b> .....	55
2.5.1. Biossíntese de L-fenilalanina .....	56
2.5.2. Via de degradação do glicerol e produção de butanol .....	58
2.5.3. Biossíntese de violaceína.....	59
<b>3. Resultados e discussão</b> .....	61
<b>3.1. Avaliação da abrangência de <i>Component Contribution</i> e <i>Reaction Fingerprinter</i></b> .....	63
3.1.1. Reações de isomerização .....	79
3.1.2. Reações que envolvem moléculas com estruturas desconhecidas .....	81
3.1.3. Reações que envolvem moléculas que não podem ser decompostas em grupos....	83
3.1.4. Outros casos particulares .....	86
3.1.5. Reações com <i>Err<sup>std</sup></i> muito elevado.....	87
3.1.6. Efeito da definição do limiar de dissemelhança em <i>Reaction Fingerprinter</i> .....	90
3.1.7. Emparelhamento de reações em sentidos opostos.....	91
<b>3.2. Avaliação da complementaridade entre <i>Component Contribution</i> e <i>Reaction Fingerprinter</i></b> .....	92
<b>3.3. Desempenho de <i>Reaction Fingerprinter</i> para reações com <math>\Delta G_R^0</math> conhecido mas ausentes de TECR</b> .....	96
<b>3.4. Análise de vias metabólicas</b> .....	99
3.4.1. Biossíntese de L-fenilalanina .....	99
3.4.2. Degradação de glicerol e produção de butanol .....	103
3.4.3. Biossíntese da violaceína.....	107
<b>4. Conclusões e trabalho futuro</b> .....	111
<b>4.1. Aperfeiçoamentos em <i>Component Contribution</i></b> .....	113
<b>4.2. Aperfeiçoamentos em <i>Reaction Fingerprinter</i></b> .....	114
<b>4.3. Extensão das bases de dados sobre a termodinâmica de reações bioquímicas</b> .	115
<b>4.4. Trabalho futuro</b> .....	117
<b>Referências</b> .....	119

## ÍNDICE DE FIGURAS

<b>Fig. 1.1.</b> - Distribuição das moléculas de um gás em diferentes volumes.	<b>7</b>
<b>Fig. 1.2.</b> - Decomposição da molécula de crotonato em grupos funcionais.	<b>16</b>
<b>Fig. 1.3.</b> - Reacções de um ciclo metabólico fútil com inconsistências na aplicação de <i>RC</i> e <i>GC</i> .	<b>20</b>
<b>Fig. 1.4.</b> - Decomposição do vetor de uma reacção em <i>RC</i> e <i>GC</i> .	<b>21</b>
<b>Fig. 1.5.</b> - Enumeração dos atributos químicos de uma molécula de piruvato de acordo com <i>IGERS</i> .	<b>23</b>
<b>Fig. 1.6.</b> - Identificação das alterações nos atributos químicos entre reagentes e produtos em <i>IGERS</i> .	<b>24</b>
<b>Fig. 1.7.</b> - Exemplo do cálculo do índice de semelhança entre duas reacções em <i>IGERS</i> .	<b>25</b>
<b>Fig. 2.1.</b> - Ilustração dos conteúdos do ficheiro MDL de uma molécula de isocitrato.	<b>45</b>
<b>Fig. 2.2.</b> - Ilustração dos conteúdos do ficheiro RXN da reacção de isomerização do isocitrato a citrato.	<b>46</b>
<b>Fig. 2.3.</b> - Identificação dos átomos e ligações químicas que são modificados na perspectiva da construção de uma <i>Reaction Fingerprint</i> .	<b>48</b>
<b>Fig. 2.4.</b> - Alargamento do centro reactivo à primeira ligação química na construção de uma <i>Reaction Fingerprint</i> .	<b>49</b>
<b>Fig. 2.5.</b> - Alargamento do centro reactivo à segunda ligação química na construção de uma <i>Reaction Fingerprint</i> .	<b>49</b>
<b>Fig. 2.6.</b> - KEGG Reaction R00019.	<b>52</b>
<b>Fig. 2.7.</b> - KEGG Reaction R00076.	<b>52</b>
<b>Fig. 2.8.</b> - KEGG Reaction R00069.	<b>53</b>
<b>Fig. 2.9.</b> - Estrutura química bidimensional do aminoácido L-fenilalanina.	<b>56</b>
<b>Fig. 2.10.</b> - Estrutura química bidimensional do butanol.	<b>58</b>
<b>Fig. 2.11.</b> - Estrutura química bidimensional da violaceína.	<b>59</b>
<b>Fig. 3.1.</b> - Representação gráfica do desempenho geral de <i>Component Contribution</i> para a amostra de 200 reacções de KEGG.	<b>69</b>
<b>Fig. 3.2.</b> - KEGG Reaction R03265.	<b>79</b>
<b>Fig. 3.3.</b> - KEGG Reaction R06380.	<b>79</b>
<b>Fig. 3.4.</b> - KEGG Reaction R09283.	<b>80</b>
<b>Fig. 3.5.</b> - KEGG Reaction R07839.	<b>80</b>
<b>Fig. 3.6.</b> - KEGG Reaction R09292.	<b>81</b>
<b>Fig. 3.7.</b> - KEGG Reaction R02718.	<b>81</b>

<b>Fig. 3.8.</b> – KEGG Reaction R10306.	<b>82</b>
<b>Fig. 3.9.</b> – KEGG Reaction R05092.	<b>84</b>
<b>Fig. 3.10.</b> – KEGG Reaction R09059.	<b>84</b>
<b>Fig. 3.11.</b> – KEGG Reaction R07726.	<b>86</b>
<b>Fig. 3.12.</b> – KEGG Reaction R09043.	<b>86</b>
<b>Fig. 3.13.</b> – KEGG Reaction R06987.	<b>91</b>
<b>Fig. 3.14.</b> – KEGG Reaction R00212.	<b>92</b>
<b>Fig. 3.15.</b> – Dispersão das estimativas geradas por <i>Component Contribution</i> e <i>ReactionFingerprinter</i> para a amostra de 200 reações de KEGG.	<b>96</b>
<b>Fig. 3.16</b> – Dispersão das estimativas geradas por <i>Component Contribution</i> e <i>ReactionFingerprinter</i> para as reações da via de biossíntese de L-fenilalanina.	<b>101</b>
<b>Fig. 3.17</b> – Dispersão das estimativas geradas por <i>Component Contribution</i> e <i>ReactionFingerprinter</i> para as reações da via de degradação do glicerol e produção de butanol.	<b>105</b>
<b>Fig. 3.18</b> – Dispersão das estimativas geradas por <i>Component Contribution</i> e <i>ReactionFingerprinter</i> para as reações da via de biossíntese de violaceína.	<b>108</b>

## ÍNDICE DE TABELAS

---

---

<b>Tabela 3.1.</b> – Reações selecionadas para comparação dos valores de $\Delta G_r^0$ estimados por <i>ReactionFingerprinter</i> com valores experimentais.	<b>55</b>
<b>Tabela 3.1.</b> – Estimativas para $\Delta G_r^0$ obtidas por <i>Component Contribution</i> para a amostra de 200 reações de KEGG.	<b>64</b>
<b>Tabela 3.2.</b> – Estimativas para $\Delta G_r^0$ obtidas por <i>ReactionFingerprinter</i> para a amostra de 200 reações de KEGG.	<b>71</b>
<b>Tabela 3.3.</b> – Quadro-resumo das reações para as quais não foi possível gerar estimativas por algum dos métodos.	<b>93</b>
<b>Tabela 3.4.</b> – Estimativas para $\Delta G_r^0$ obtidas por <i>ReactionFingerprinter</i> para a amostra de 10 reações extra-TECR.	<b>98</b>
<b>Tabela 3.5.</b> – Estimativas para $\Delta G_r^0$ obtidas por <i>Component Contribution</i> para as reações da via de biossíntese de L-fenilalanina.	<b>100</b>
<b>Tabela 3.6.</b> – Estimativas para $\Delta G_r^0$ obtidas por <i>ReactionFingerprinter</i> para as reações da via de biossíntese de L-fenilalanina.	<b>102</b>
<b>Tabela 3.7.</b> – Estimativas para $\Delta G_r^0$ obtidas por <i>Component Contribtuion</i> para as reações da via de degradação do glicerol e produção de butanol.	<b>104</b>
<b>Tabela 3.8.</b> – Estimativas para $\Delta G_r^0$ obtidas por <i>ReactionFingerprinter</i> para as reações da via de degradação do glicerol e produção de butanol.	<b>106</b>
<b>Tabela 3.9.</b> – Estimativas para $\Delta G_r^0$ obtidas por <i>Component Contribtuion</i> para as reações da via de biossíntese de violaceína.	<b>107</b>
<b>Tabela 3.10.</b> – Estimativas para $\Delta G_r^0$ obtidas por <i>ReactionFingerprinter</i> para as reações da via de biossíntese de violaceína.	<b>108</b>

---



## 1. INTRODUÇÃO

---



Desde a década de 1970 que o conhecimento acumulado sobre o metabolismo celular, aliado ao advento das técnicas de engenharia e expressão genética, tem conferido um caráter cada vez mais real e alargado à aplicação de microrganismos como bio-reatores de elevada especificidade e alto rendimento.<sup>[1]</sup> Esta aplicação de âmbito alargado tem uma relevância cada vez maior no mundo da biotecnologia, manifestando-se tanto como uma ferramenta indispensável em domínios concretos da investigação científica, como em conhecidos casos de sucesso nas mais diversas áreas da indústria. São disso exemplo o recurso generalizado às mais variadas estirpes de *Escherichia coli*<sup>[2]</sup> e *Saccharomyces cerevisiae*<sup>[3]</sup> enquanto modelos de referência na investigação ligada à biotecnologia, bem como a aplicação orientada de microrganismos em vários ramos da indústria química, alimentar e farmacêutica<sup>[4, 5]</sup>.

No entanto, os recursos e conhecimentos já disponíveis nestas áreas da indústria e da investigação geram também a necessidade de compreender e avaliar quantidades avassaladoras de informação que, frequentemente, continuam a carecer de métodos de análise eficientes.<sup>[6, 7, 8, 9]</sup> Por exemplo, a obtenção de um produto específico, obtido como resultado da atividade metabólica de um determinado microrganismo, pode ser conseguida através de várias vias: em primeiro lugar, a escolha do próprio microrganismo a utilizar pode recair sobre um em particular, entre diversas possibilidades distintas<sup>[10, 11]</sup>; além disso, a obtenção do produto desejado pode ser conseguida em circunstâncias muito diversificadas – entre outras, podem enumerar-se a composição do substrato a utilizar, com efeito nos recursos que são disponibilizados ao organismo, ou as vias metabólicas envolvidas, que poderão ser alvo das mais variadas formas de estimulação e/ou supressão, consoante o objectivo traçado.<sup>[12, 13]</sup> Enumerar as imensas possibilidades envolvidas em tal processo é um exercício complexo, que rapidamente demonstra por si só a importância de criar métodos capazes de processar elevadas quantidades de informação. As ferramentas computacionais e a bioinformática adquirem aqui um papel crucial, com o potencial de proporcionar a resposta adequada a uma grande parte das limitações ainda existentes.<sup>[6, 7, 8, 9, 14]</sup>

Além de ferramentas capazes de coligir e apresentar grandes quantidades de informação, frequentemente é ainda necessário encontrar formas de comparação eficazes e que funcionem como um verdadeiro fator de apoio à decisão na biotecnologia. Neste contexto, o conhecimento rigoroso acerca da termodinâmica das reações

## **Introdução**

bioquímicas e, conseqüentemente, das vias metabólicas em que se inserem, constitui uma ferramenta valiosa no que diz respeito à previsão da direcionalidade que deverá caracterizá-las<sup>[15, 16]</sup>, principalmente quando avaliadas sob condições específicas como as que se verificam no interior das células de organismos vivos. A energia livre de Gibbs insere-se nesta problemática como o parâmetro mais relevante a ter em conta no que toca à previsão da direcionalidade de uma reação química. Em última análise, a obtenção de medições ou estimativas com elevado grau de precisão para este parâmetro termodinâmico poderá servir como um poderoso fator de decisão na avaliação e seleção das vias metabólicas mais eficientes e cuja exploração seja mais promissora.<sup>[15, 16, 17, 18, 19, 20]</sup>

Definindo-se, de um ponto de vista prático, e em termos muito simplificados, como o saldo energético de uma reação química, a variação da energia livre de Gibbs indica se uma dada reação ocorre com consumo ou libertação de energia. A ocorrência de uma reação com libertação de energia (correspondente a um valor negativo na variação da energia livre de Gibbs) é sinónima de que essa reação ocorre de forma espontânea e é, portanto, favorecida nas condições sob as quais esteja a ser avaliada.<sup>[21]</sup> Finalmente, a comparação da energia livre de Gibbs envolvida em reações diferentes, ou mesmo vias metabólicas completas, permite diferenciar numerosos casos de maior ou menor potencial para aplicação com vista a um objectivo específico.<sup>[16, 17, 18, 19, 20]</sup> Reconhecer a força deste parâmetro termodinâmico como um importante fator de apoio à decisão em biotecnologia implica reconhecer também a preponderância da obtenção de estimativas válidas para a sua magnitude.

Porém, existem limitações práticas à determinação da energia livre de Gibbs por métodos experimentais e as ferramentas computacionais disponíveis baseiam-se em metodologias que divergem entre si em vários aspetos, com reflexo ao nível dos resultados que produzem. Conseqüentemente, as estimativas produzidas por estas ferramentas revelam ainda espaço para melhoria, nomeadamente ao nível do seu desfasamento em relação à realidade, do número e variedade de reações a que podem ser aplicadas com sucesso, e mesmo do tempo requerido para computação.<sup>[17, 20]</sup>

Surge assim a motivação para a presente dissertação, através da qual se pretende fazer uma análise detalhada sobre alguns dos métodos computacionais existentes para gerar estimativas da variação da energia livre de Gibbs em reações bioquímicas, expondo os respetivos méritos e limitações. Com esta análise, espera-se ainda encontrar

e enumerar possíveis formas de complementaridade entre os métodos abordados, traçando assim as bases para a criação de uma ferramenta mais abrangente e mais poderosa. Ao longo das próximas secções serão especificados pormenores sobre a temática abordada, proporcionando uma melhor compreensão global do problema e dos objectivos do presente trabalho.

## 1.1. A TERMODINÂMICA DAS REAÇÕES BIOQUÍMICAS E A ENERGIA LIVRE DE GIBBS<sup>[21]</sup>

---

Apesar de as reações bioquímicas ocorrerem sob condições muito particulares e, por vezes, difíceis de analisar à luz da termodinâmica, todos os fenómenos biológicos obedecem às suas leis. A compreensão destes fenómenos de um ponto de vista termodinâmico é importante ao nível da descrição dos processos metabólicos em termos quantitativos, bem como enquanto uma forma de prever se estes podem mesmo ocorrer – ou seja, se são processos espontâneos.

Em termodinâmica, o sistema é definido como a parte do universo que é relevante para o processo ou reacção sob análise. O resto do universo é designado como a vizinhança. No caso das reações bioquímicas, habitualmente a célula é considerada como o sistema sob estudo.

---

### 1.1.1. A PRIMEIRA LEI DA TERMODINÂMICA

---

No sistema existe uma determinada quantidade de energia, habitualmente simbolizada pela letra  $U$ , e, pela primeira lei da termodinâmica, é estabelecido que existe conservação de energia no universo, ou seja, a energia não pode ser criada nem destruída. Contudo, a energia pode ser transferida, nomeadamente quando ocorre uma transformação no sistema. Durante essa transformação, parte da energia pode ser utilizada para produzir trabalho. Neste processo, a variação de energia no sistema é definida como a diferença entre o calor,  $q$ , absorvido pelo sistema a partir da vizinhança e o trabalho,  $w$ , realizado pelo sistema sobre a vizinhança. O carater grego  $\Delta$  é

## **Introdução**

normalmente utilizado para designar a variação.

$$\Delta U = U_{final} - U_{inicial} = q - w \quad (\text{Equação 1.1.})$$

A energia do sistema pode ser usada para a realização de variadíssimas formas de trabalho, como é o caso da expansão de um gás com o aumento da temperatura. É também frequente haver referências a diversas formas de energia – mecânica, química, eléctrica – sendo que praticamente todas podem ser relevantes no estudo de sistemas biológicos.

É, no entanto, importante notar que a grande maioria dos processos biológicos ocorrem sob pressão constante. Neste caso, o trabalho realizado pela expansão de um gás é definido como o produto entre a magnitude da pressão no sistema e a variação ocorrida no volume do gás –  $P\Delta V$ . Assim, pode então definir-se uma outra grandeza termodinâmica: entalpia,  $H$ .

$$H = U + PV \quad (\text{Equação 1.2.})$$

Quando o sistema sofre uma transformação sob pressão constante, temos:

$$\Delta H = \Delta U + P\Delta V = q_p - w + P\Delta V \quad (\text{Equação 1.3.})$$

sendo que  $q_p$  é definido como o calor a pressão constante. Sabendo ainda que neste sistema  $w = P\Delta V$ ,

$$\Delta H = q_p - P\Delta V + P\Delta V = q_p \quad (\text{Equação 1.4.})$$

Por outras palavras, a variação na entalpia corresponde ao calor sob pressão constante. Além disso, é válido dizer que para a grande maioria das reacções bioquímicas, as variações no volume do sistema são praticamente inexistentes, pelo que a variação energética no sistema pode ser equiparada à variação da entalpia. À semelhança da energia, do calor e do trabalho, a entalpia é medida em joule ( $J$ ).

---

### 1.1.2. SEGUNDA LEI DA TERMODINÂMICA

---

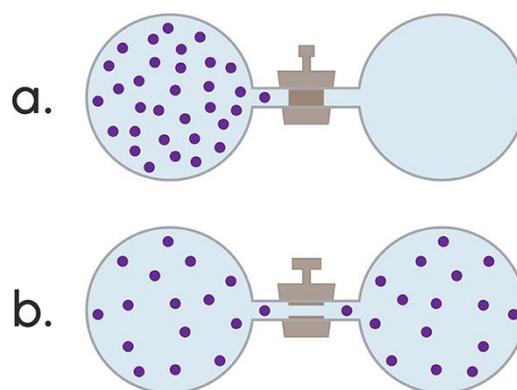
Como referido anteriormente, a termodinâmica permite prever a espontaneidade

de um processo. Um processo é designado espontâneo quando ocorre sem o fornecimento de energia da vizinhança ao sistema. Contudo, a primeira lei da termodinâmica, por si só, não permite estabelecer se um processo, ou reação, é ou não espontâneo. Por exemplo, o calor é transferido de forma espontânea de um corpo a uma temperatura elevada para outro a temperatura mais baixa, mas nunca ao contrário. Porém, qualquer dos casos estaria de acordo com a primeira lei da termodinâmica, visto que a energia total envolvida neste processo nunca é alterada.

De acordo com a segunda lei da termodinâmica, um processo espontâneo está relacionado com uma alteração no grau de desordem do sistema. Note-se que, neste contexto, o termo desordem refere-se ao número de formas energeticamente equivalentes entre si de organizar espacialmente os componentes do sistema,  $W$ . A distribuição das moléculas de um gás num volume maior ou menor é um exemplo disso, como ilustra a Fig. 1.1. O grau de desordem, ou aleatoriedade, no sistema é quantificado pela entropia,  $S$ :

$$S = k_B \ln W \quad (\text{Equação 1.5.})$$

em que  $k_B$  corresponde à constante de Boltzmann. A entropia é expressa em  $J \cdot K^{-1}$  e o arranjo mais provável do sistema será sempre aquele que maximiza  $W$  e, conseqüentemente,  $S$ .



**Fig. 1.1.** – A figura representa as moléculas de um gás, contidas em duas ampolas de igual tamanho, unidas por uma válvula. Em a), a referida válvula está fechada e o gás está inteiramente contido na ampola do lado esquerdo, caracterizando-se o sistema por um baixo valor de entropia. Ao abrir a válvula, como em b), a entropia do sistema aumenta à medida que as moléculas do gás se difundem para a ampola do lado direito, até que a quantidade de gás em cada ampola é igualada.

Assim, se um processo espontâneo como o ilustrado na Fig. 1.1. é caracterizado por

## Introdução

variações nulas de energia e entalpia ( $\Delta U = 0$  e  $\Delta H = 0$ ), a variação da entropia deverá ser sempre maior do que zero ( $\Delta S > 0$ ). Por outras palavras, o número de arranjos energeticamente equivalentes dos componentes do sistema no seu estado final deverá ser maior do que o número de arranjos no estado inicial. De resto, todos os processos espontâneos tendem a aumentar a entropia do universo:

$$\Delta S_{\text{sistema}} + \Delta S_{\text{vizinhança}} = \Delta S_{\text{Universo}} > 0 \text{ (Equação 1.6.)}$$

No caso dos processos químicos e biológicos é, de um ponto de vista prático, impossível determinar a entropia do sistema através da medição de  $W$ . Contudo, existe uma expressão equivalente para a entropia, aplicável a processos que decorrem a temperatura constante, como, de resto, é o caso da grande maioria das reações metabólicas:

$$\Delta S \geq \frac{q}{T} \quad \text{(Equação 1.7.)}$$

Assim, em processos espontâneos, a variação na entropia pode ser determinada a partir da medição do calor envolvido numa reação.

---

### 1.1.3. ENERGIA LIVRE DE GIBBS E PROCESSOS ESPONTÂNEOS

---

Mesmo sabendo que a entropia do Universo tende a aumentar, a espontaneidade de um processo não pode ser determinada apenas através da variação de entropia que o caracteriza. Por exemplo, a 273 K, a congelação da água é um processo espontâneo. Contudo, a formação de gelo obriga ao constrangimento das moléculas da água, num arranjo espacial que claramente diminui a entropia face ao arranjo que se verifica quando a água está no estado líquido. É então necessário definir um outro critério para que um determinado processo possa ser considerado espontâneo.

As *Equação 1.4.* a *Equação 1.7.* permitem estabelecer que, sob pressão e temperatura constantes:

$$\Delta S \geq \frac{q_p}{T} = \frac{\Delta H}{T} \quad \text{(Equação 1.8.)}$$

pelo que:

$$\Delta H - T\Delta S \leq 0 \quad (\text{Equação 1.9.})$$

A *Equação 1.9.* é o verdadeiro critério para determinar a espontaneidade de uma reação. Foi definido em 1878 por J. Willard Gibbs, e estabelece a energia livre de Gibbs,  $G$ , como:

$$G = H - TS \quad (\text{Equação 1.10.})$$

A variação da energia livre de Gibbs de um processo é  $\Delta G$ . Consequentemente, os processos espontâneos ocorridos sob condições de pressão e temperatura constantes são caracterizados por:

$$\Delta G = \Delta H - T\Delta S < 0 \quad (\text{Equação 1.11.})$$

Estes processos são também designados como exergónicos. Os processos não espontâneos são conhecidos como endergónicos e são caracterizados por valores positivos de  $\Delta G$  – necessitam do fornecimento de energia livre da vizinhança para que possam ocorrer. Se um determinado processo é exergónico, então o processo inverso será endergónico, e vice-versa. Assim sendo, o valor de  $\Delta G$  de um processo ou reacção indica se o mesmo pode acontecer espontaneamente no sentido descrito. No caso de processos em equilíbrio – aqueles em que existe um balanceamento entre os sentidos direto e inverso –  $\Delta G$  é igual a zero. Para a grande maioria das reacções, não é possível fazer medições do valor absoluto de  $G$ ,  $H$  e  $S$ ; no entanto, a sua variação ao longo do processo pode, frequentemente, ser medida ou estimada –  $\Delta G$ ,  $\Delta H$  e  $\Delta S$ .

Um processo acompanhado por um aumento da entalpia ( $\Delta H > 0$ ) pode ocorrer espontaneamente caso a variação na entropia seja suficientemente positiva ( $\Delta S > 0$ ). Da mesma forma, um processo acompanhado por uma diminuição da entropia ( $\Delta S < 0$ ) necessita que a respectiva variação da entalpia seja suficientemente negativa ( $\Delta H < 0$ ). É ainda importante enfatizar que um valor de  $\Delta G$  muito negativo não assegura que uma reação química ou bioquímica ocorra a um ritmo mensurável. Este ritmo, ou a taxa de conversão de um substrato num determinado produto depende sempre da existência de um mecanismo de reação, que por sua vez é independente de  $\Delta G$ . No caso particular das

## Introdução

reações metabólicas, esse mecanismo é normalmente assegurado pela enzima que cataliza o processo e que assim permite acelerar a velocidade a que o mesmo ocorre.

A energia livre de Gibbs, bem como a entalpia e a entropia são funções de estado. Tal significa que o seu valor depende exclusivamente do estado ou propriedades atuais do sistema, e não da maneira como esse estado foi alcançado. Neste sentido, as medições ou estimativas dos parâmetros termodinâmicos de um processo são levadas a cabo através da consideração apenas do estado inicial e final do sistema, e ignorando todas as variações intermédias que possam ter ocorrido na entalpia e na entropia.

---

### 1.1.4. VARIAÇÃO DA ENERGIA LIVRE DE GIBBS NO ESTADO DE EQUILÍBRIO

---

A entropia de uma substância aumenta proporcionalmente com o aumento do volume ao qual está confinada. Por exemplo, as moléculas de uma determinada substância, quando dissolvidas num líquido, tendem a distribuir-se uniformemente por todo o seu volume. Assim sendo, a entropia pode ser avaliada em função da concentração.

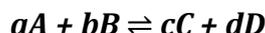
Havendo uma variação na entropia em função da concentração, então também a energia livre é afetada. Consequentemente, a variação na energia livre de uma reação depende da concentração tanto dos reagentes como dos produtos. No contexto deste trabalho, esta última assumpção assume especial importância, uma vez que muitas reações bioquímicas podem ocorrer de forma espontânea, e em ambos os sentidos, dependendo da concentração relativa dos seus reagentes e produtos.

A relação entre a concentração e a energia livre de uma substância  $A$  é aproximadamente

$$\bar{G}_A = \bar{G}_A^{\circ} + RT \ln a_A \quad (\text{Equação 1.12.})$$

onde  $\bar{G}_A$  corresponde ao potencial químico (ou energia livre parcial molar) de  $A$  (a barra transversal sobre  $G$  indica a quantidade por mole),  $\bar{G}_A^{\circ}$  é o potencial químico de  $A$  em condições padrão (ver secção 1.1.5.),  $R$  é a constante dos gases ideais e  $a_A$  a atividade de  $A$ . A atividade é dada pela fugacidade de  $A$ ,  $f_A$ , multiplicada pela concentração molar de

$A$ ,  $[A]$ . Para soluções diluídas,  $f$  toma normalmente o valor de  $1 M^{-1}$ , pelo que frequentemente pode assumir-se que  $a_A = [A]$ . Assim sendo, para a reação genérica:



a variação na energia livre do processo no sentido direto será:

$$\Delta G = c\bar{G}_C + d\bar{G}_D - a\bar{G}_A - b\bar{G}_B \quad (\text{Equação 1.13.})$$

e

$$\Delta \bar{G}^o = c\bar{G}_C^o + d\bar{G}_D^o - a\bar{G}_A^o - b\bar{G}_B^o \quad (\text{Equação 1.14.}),$$

uma vez que as energias livres de cada reagente ou produto são aditivas, e que a variação da energia livre de uma reação corresponde ao somatório das energias livres dos produtos menos as dos reagentes. Usando, para cada substância, relações do tipo das representadas na *Equação 1.12.*, temos:

$$\Delta G = \Delta G^o + RT \ln \left( \frac{[C]^c [D]^d}{[A]^a [B]^b} \right) \quad (\text{Equação 1.15.})$$

onde  $\Delta G^o$  corresponde à variação da energia livre da reação, quando todos os seus reagentes e produtos se encontram em condições padrão. Assim, a expressão para a variação da energia livre de uma reação pode ser decomposta em duas partes: (1) um termo constante que depende unicamente da própria reação e (2) um termo variável, dependente da concentração de reagentes e produtos, da estequiometria da reação e da temperatura.

Numa reação no estado de equilíbrio esta variação não existe, visto que a variação a energia livre dos reagentes é igual à energia livre dos produtos. Em resultado disso,  $\Delta G = 0$ , pelo que a *Equação 1.15.* fica:

$$\Delta G^o = -RT \ln K_{eq} \quad (\text{Equação 1.16.})$$

onde  $K_{eq}$  corresponde à constante de equilíbrio da reação:

## Introdução

$$K_{eq} = \frac{C_{eq}^c D_{eq}^d}{A_{eq}^a B_{eq}^b} = e^{-\Delta G^0/RT} \quad (\text{Equação 1.17.})$$

A abreviatura em índice “*eq*” indica que se trata das concentrações de reagentes e produtos no estado de equilíbrio. (Neste contexto, a condição de equilíbrio é normalmente assumida por defeito, pelo que é frequente a omissão desta abreviatura.)

A forma como a constante de equilíbrio varia com a temperatura pode também ser expressa através da substituição dos termos da *Equação 1.16.* pelos da *Equação 1.11.:*

$$\ln K_{eq} = \frac{-\Delta H^0}{R} \left(\frac{1}{T}\right) + \frac{\Delta S^0}{R} \quad (\text{Equação 1.18.})$$

onde  $H^0$  e  $S^0$  representam, respetivamente, entalpia e entropia em condições padrão. O gráfico de  $\ln K_{eq}$  em função de  $1/T$ , conhecido como gráfico de van't Hoff, possibilita a determinação de  $\Delta H^0$  e  $\Delta S^0$  (e, conseqüentemente,  $\Delta G^0$ ) a partir da medição de  $K_{eq}$  sob duas, ou mais, temperaturas diferentes. Habitualmente, este é um método mais prático do que a medição de  $\Delta H$  e  $\Delta S$  por calorimetria (medição do calor,  $q_p$ , de um processo).

---

### 1.1.5. CONVENÇÕES SOBRE CONDIÇÕES-PADRÃO

---

De modo a que se possam comparar variações na energia livre de Gibbs de diferentes reações, foi necessário que  $\Delta G$  fosse expressa relativamente a um estado padrão. De acordo com a convenção seguida em química física, um determinado soluto encontra-se em condições padrão quando a uma temperatura de  $25^\circ C$ , uma pressão de  $1 atm$  e a sua atividade equivale a 1 (concentração corrigida para um comportamento não ideal, para concentrações maiores do que numa diluição infinita). As reações metabólicas são um dos casos em que a respetiva atividade pode ser equiparada à sua concentração molar (frequentemente situada na ordem dos milimolar ou inferior). Contudo, e porque estas reações ocorrem normalmente a pH praticamente neutro, em bioquímica adotaram-se condições padrão diferentes das enumeradas anteriormente:

- 1) Atribui-se o valor 1 à atividade da água pura, ainda que a sua concentração seja de  $55,5 M$ . Esta prática simplifica substancialmente a expressão da energia livre de reações que envolvem soluções diluídas nas quais a água actua como um dos

reagentes, uma vez que permite ignorar o termo  $[H_2O]$ .

- 2) É também atribuído o valor 1 à atividade do ião hidrogénio ( $H^+$ ), quando se consideram reações sob as condições fisiológicas em que o pH é 7. Em bioquímica, assume-se por isso que o valor do pH em condições padrão é 7,0 (pH neutro, com  $[H^+] = 10^{-7} M$ ), em vez de 0 ( $[H^+] = 1 M$ ) – condição na qual a maioria das moléculas biológicas é instável.
- 3) O estado padrão de uma substância capaz de intervir numa reação ácido-base é definido em função da concentração total da mistura de todos os seus estados de ionização sob pH 7. Ao invés, em química física a convenção prevalente assume uma espécie pura, independentemente de esta realmente ocorrer a pH 0. A vantagem desta convenção bioquímica é que a concentração total de uma substância com vários estados de ionização – como é o caso da maioria das moléculas biológicas – é normalmente mais fácil de medir do que a concentração de cada um dos seus estados ionizados. No entanto, e uma vez que esta composição iónica de um ácido ou de uma base varia com o pH, as energias livres padrão calculadas de acordo com a convenção bioquímica são válidas unicamente para pH 7.

Sob a convenção da bioquímica, as variações padrão da energia livre são frequentemente simbolizadas por  $\Delta G^{\circ'}$ , de modo a que possam ser diferenciadas daquelas que são calculadas sob a convenção da química física ( $\Delta G^{\circ}$ ). (Também aqui, quando o contexto é unicamente bioquímico, é habitual este pormenor ser omitido.) Note-se também que numa reacção que não envolva  $H_2O$ ,  $H^+$  ou espécies ionizáveis,  $\Delta G^{\circ'} = \Delta G^{\circ}$ .

## 1.2. ESTIMATIVA DA ENERGIA LIVRE DE GIBBS POR MÉTODOS COMPUTACIONAIS

---

O número de reações bioquímicas para as quais se conhece o valor da variação na energia livre de Gibbs é muito limitado e não chega, nem de perto, para cobrir a vastidão de reações e vias metabólicas já descritas.<sup>[17, 20, 22]</sup> Além disto, os métodos tradicionais

## **Introdução**

que permitem a determinação destes dados envolvem muitas horas de trabalho e um elevado consumo de recursos: implicam a medição direta da constante de equilíbrio aparente,  $K'$ , de cada reação, seguida da aplicação da *Equação 1.16.* Para tal, é preparada uma solução tampão, à qual é(são) adicionado(s) o(s) substrato(s) da reação, juntamente com uma enzima capaz de catalisar a transformação a ser analisada. Atingido o estado de equilíbrio – habitualmente determinado por leituras da densidade ótica -, calcula-se o quociente da reação,  $Q$ , através da divisão das concentrações de produtos pelas concentrações de substratos. É também recomendável a execução do processo inverso, partindo do que eram antes os produtos e medindo a formação de substrato do processo inicial. Atingido um estado de equilíbrio, os valores de  $Q$  deverão ser semelhantes entre si e equiparáveis ao valor de  $K'$ .<sup>[23, 24, 25, 26, 27, 28, 29]</sup> Todo este processo exige a purificação de grandes quantidades da enzima catalisadora de cada processo, e ainda a execução de um elevado número de medições. É também de salientar que este método é limitado pela capacidade instrumental de detetar a evolução nas concentrações de substratos e produtos, pelo que nem sempre pode ser aplicado com a precisão desejada (por exemplo, reações cuja  $\Delta G^\circ$  esteja afastada de zero por muitas ordens de grandeza tornam-se muito difíceis de analisar por este método com uma precisão razoável, uma vez que os reagentes tendem a ser consumidos praticamente na sua totalidade e que o atual estado da arte dos instrumentos utilizados não permite a medição de concentrações tão próximas de zero). Tudo somado, e atendendo aos milhares de reações que é necessário analisar, a via experimental torna-se muito dispendiosa e pouco suscetível de, em tempo útil, cobrir as grandes lacunas existentes nesta área.<sup>[17]</sup>

Por todos estes motivos, e sendo bem conhecidos os fatores que influenciam a variação da energia livre de Gibbs, os métodos computacionais capazes de produzir estimativas sólidas para este parâmetro são alvo de um interesse cada vez maior.<sup>[17, 20]</sup> Foram já descritas diversas abordagens, resultando numa maior ou menor aceitação, consoante a validade dos resultados produzidos e a viabilidade da sua aplicação prática e em larga escala. Ao longo dos próximos parágrafos será feita uma breve descrição dos exemplos mais relevantes de abordagens e ferramentas já existentes, sublinhando sempre as suas principais vantagens e limitações. O balanço efetuado permitirá definir o rumo a seguir nesta dissertação e o estabelecimento de objetivos concretos.

---

**1.2.1. CONTRIBUIÇÕES DE GRUPOS FUNCIONAIS – ‘GROUP CONTRIBUTION’**

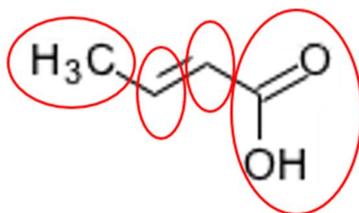
---

O método das ‘*group contributions*’ tem por base a utilização da energia livre de formação,  $\Delta G_f$ , de cada um dos grupos funcionais que constituem um determinado composto químico para o cálculo da energia livre de formação desse composto.<sup>[20]</sup> Por outras palavras, assume que  $\Delta G_f$  desse composto é equiparável à soma de  $\Delta G_f$  de cada um dos diferentes grupos que o constituem. Por exemplo, considere-se um composto orgânico do tipo  $\mathbf{AB}_2\text{-AB}_2$ , para o qual o valor de  $\Delta G_f$  é desconhecido. Caso se conheça – ou seja possível estimar – o valor de  $\Delta G_f$  de  $\mathbf{AB}_2$ , então  $\Delta G_f$  do composto pode ser estimado através da soma de  $\Delta G_f$  dos dois grupos  $\mathbf{AB}_2$ . Embora este exemplo esteja extremamente simplificado, é suficiente para explicitar o princípio básico sobre o qual assenta o método das contribuições de grupos funcionais: a energia livre de formação de um dado composto é estimada através das contribuições individuais de cada um dos grupos químicos que o constituem. Consequentemente, o cálculo da energia livre de Gibbs de uma reação química,  $\Delta G_r$ , deixa de estar centrado nas energias livres de formação de produtos e reagentes (*Equação 1.14.*): passa a ser calculada pela soma dos  $\Delta G_f$  dos grupos funcionais produzidos pela reação, à qual é subtraída a soma dos  $\Delta G_f$  daqueles que são consumidos.

A possibilidade de implementar uma abordagem deste género foi primeiramente descrita por S. Benson e J. Buss<sup>[30]</sup>, cujo trabalho se debruçou sobre a aditividade de propriedades intramoleculares enquanto forma de estimar as propriedades termodinâmicas de diferentes substâncias. Nesta abordagem, os autores definiram as fronteiras de um “grupo” como pares de átomos ou elementos estruturais contidos dentro de um raio de 3 a 5 Å. Este método foi inicialmente aplicado com relativo sucesso para estimar os parâmetros termodinâmicos de gases ideais<sup>[30]</sup> e posteriormente alargado ao estudo de fases líquidas e sólidas<sup>[31]</sup>. Relativamente ao trabalho de S. Benson e J. Buss, a aplicação das contribuições de grupos funcionais às reacções bioquímicas surgiu bastante mais tarde<sup>[32, 33]</sup>, mas rapidamente se tornou no método mais utilizado para o cálculo de energias livres de Gibbs nesse contexto específico<sup>[18, 34]</sup>. A aplicação da decomposição em grupos de uma molécula em contexto real pode ser exemplificada através do caso do crotonato<sup>[32]</sup>, um ácido orgânico com a fórmula  $\text{CH}_3\text{-CH=CH-COO}^-$ : a sua decomposição em  $-\text{CH}_3$ ,  $-\text{CH=}$  ( $x_2$ ) e  $\text{COO}^-$ , de acordo com o ilustrado na **Fig. 1.2.**

## Introdução

permite estimar a respetiva energia livre de formação em  $-279,1 \text{ kJ/mol}$ , um valor muito próximo dos  $-277,4 \text{ kJ/mol}$  tabelados originalmente.



**Fig. 1.2.** – Decomposição do crotonato para determinação da sua energia livre de formação através das contribuições de grupos funcionais. Cada elipse a vermelho circunda um dos grupos resultantes do processo de decomposição.

No que ao valor da energia livre de Gibbs diz respeito, a grande vantagem do método *Group Contribution (GC)* está no extenso número de reacções metabólicas para as quais é capaz de gerar estimativas, incluindo praticamente todas as reacções bioquímicas cientificamente relevantes. Por exemplo, esta abordagem é capaz de cobrir cerca 70 % das reacções do metabolismo humano<sup>[19]</sup>; no caso da *E. coli*, este número sobe para perto dos 90 %<sup>[35]</sup>. Contudo, a grande limitação de *GC* reside na falta de exatidão que acaba por revelar, uma vez que parte do pressuposto de que as contribuições dos grupos definidos são puramente aditivas. O resultado é que, em média, as estimativas geradas chegam a rondar um erro entre 9 e 10  $\text{kJ/mol}$ .<sup>[34]</sup>

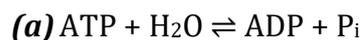
---

### 1.2.2. CONTRIBUIÇÕES DE REAGENTES – ‘*REACTANT CONTRIBUTION*’

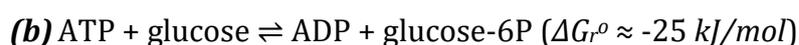
---

Pouco tempo antes da publicação do trabalho de S. Benson e J. Buss, K. Burton sugeriu que as constantes de equilíbrio aparentes,  $K'$ , juntamente com os valores experimentais da energia livre padrão de formação,  $\Delta G_f^\circ$ , de alguns compostos simples, poderiam ser utilizadas para estimar as constantes de equilíbrio de reacções para as quais aquela é desconhecida.<sup>[36]</sup> O método assim definido baseia-se na assumption de que, conhecendo-se a energia livre padrão de duas reacções diferentes,  $\Delta G_r^\circ$ , é possível calcular este parâmetro para a reacção combinada entre aquelas duas, através da soma das suas  $\Delta G_r^\circ$  individuais.

A reacção de hidrólise da ATP, (**a**), pode ser usada como um bom exemplo para explicar a lógica deste método:



Esta reacção é extremamente exergónica e ultrapassa um estado de equilíbrio que possa ser devidamente analisado pela via experimental. No entanto, as reacções catalisadas pelas enzimas glucose cinase, **(b)**, e glucose-6P fosfatase, **(c)**, ambas próximas do estado de equilíbrio, podem mais facilmente ser caracterizadas em termos de  $K'$ :



Como se pode ver, a combinação das reacções **(b)** e **(c)** dá origem a uma reacção semelhante à reacção **(a)**:



A soma de  $\Delta G_r^\circ$  das reacções **(b)** e **(c)** permite assim estimar  $\Delta G_r^\circ$  da reacção **(a)** num valor próximo de  $-37 \text{ kJ/mol}$ . Em condições padrão, assumindo concentrações no estado de equilíbrio de  $1 \text{ M}$  para ATP, ADP e  $\text{P}_i$  (e  $55 \text{ M}$  para  $\text{H}_2\text{O}$ ), o mesmo parâmetro pode ser situado entre  $-28$  e  $-34 \text{ kJ/mol}$ , dependendo da concentração do catião  $\text{Mg}^{2+}$  no meio.

De forma a tornar mais prático o cálculo dos valores de  $K'$ , e, conseqüentemente, das energias livres de Gibbs associadas a reacções bioquímicas, K. Burton publicou ainda uma extensa lista de valores de  $\Delta G_r^\circ$  de diferentes espécies químicas.<sup>[36]</sup> Parte destes valores foram citados a partir de tabelas com dados termodinâmicos, com os restantes a serem obtidos pela via aritmética da combinação de reacções. Por exemplo, considere-se o caso de uma reacção para a qual se conhecem os valores de  $\Delta G_r^\circ$  de todas as espécies intervenientes, exceto uma; considere-se ainda que o valor de  $\Delta G_r^\circ$  desta reacção pode ser obtido experimentalmente;  $\Delta G_r^\circ$  para a espécie química em falta podia então ser calculado e adicionado à tabela em questão. A existência de uma tabela deste tipo permite que  $\Delta G_r^\circ$  seja calculado para virtualmente qualquer reacção que envolva apenas espécies ou compostos presentes nessa mesma tabela.

Em 2013, Noor *et al* designaram este método por *Reactant Contribution (RC)*<sup>[20]</sup>.

## Introdução

Por uma questão prática, e pela relevância e influência exercidas pelo trabalho destes autores nesta dissertação, esta designação será também aqui adotada.

De referir ainda que, nas décadas subsequentes ao trabalho de K. Burton, originalmente publicado em 1957, vários outros autores foram publicando tabelas de energias livres de formação, mais alargadas e mais completas. Entre estas, destacam-se em particular a tabela de R. Thauer<sup>[37]</sup> e R. Alberty<sup>[38, 39]</sup>. Contudo, mesmo com todo o conhecimento acumulado sobre este tema, estima-se que apenas cerca de 10 % das reacções compreendidas pelos modelos de *E. coli* possam ser cobertas pelo método *RC*.<sup>[20]</sup>

---

### 1.2.3. CONTRIBUIÇÕES DE COMPONENTES - 'COMPONENT CONTRIBUTION'

---

O elevado erro médio inerente a *GC* e a limitada cobertura proporcionada por *RC* motivaram Noor *et al* a desenvolver um método inovador, procurando conciliar as vantagens de cada um deles de uma forma consistente. Esta nova abordagem levou à criação de *Component Contribution (CC)*, capaz de aliar a extensa cobertura de *Group Contribution* ao maior acerto do *Reactant Contribution*.<sup>[20]</sup>

Habitualmente,  $\Delta G_f^\circ$  de substâncias puras elementares no seu estado padrão é definida como sendo zero. (Por exemplo:  $O_2$ , a 298,15 K e 1 atm.) Para todas as outras espécies químicas, o mesmo parâmetro é calculado relativamente a estes pontos de referência. Esta abordagem permite determinar de forma perfeitamente consistente  $\Delta G_r^\circ$  para qualquer reacção estequiometricamente equilibrada. Contudo, quando se trata da análise termodinâmica de reacções metabólicas, pode ser importante estabelecer mais alguns pontos de referência. Em particular, existem co-fatores de elevada complexidade para os quais se torna particularmente difícil calcular ou estimar  $\Delta G_f^\circ$ . Por isso, em anteriores abordagens ao cálculo da energia livre de Gibbs em reacções bioquímicas foi vantajoso equiparar algumas espécies químicas não-elementares aos já referidos pontos de referência, para os quais  $\Delta G_f^\circ$  é considerada igual a zero. Contudo, em alguns casos concretos, esta forma de pensar cria conflitos que resultam em elevados erros no cálculo de  $\Delta G_r^\circ$ .

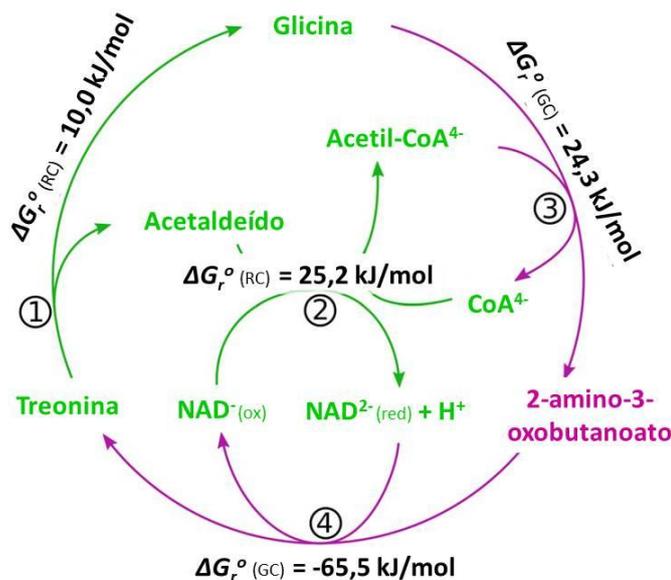
Na maioria das reacções bioquímicas em que os co-fatores mais complexos intervêm,

estes, na realidade, são usados como substratos que, estando presentes dos dois lados da equação, podem ser ignorados (à imagem do que foi feito na reação combinada da secção 1.2.2.) e a sua influência em  $\Delta G_r^o$  acaba por ser negligenciável. Por exemplo, no caso de FAD – '*flavin adenine dinucleotide*' – existe uma forma oxidada, FAD(ox)<sup>-2</sup>, que pode ser emparelhada com a correspondente forma reduzida, FAD(red)<sup>-2</sup>, com  $\Delta G_f^o$  igual a -38,88 kJ/mol, de acordo com a tabela de R. Alberty. Não obstante, existem também reações nas quais este tipo de emparelhamento não existe, como é o caso que se segue:



Nesta reação FAD(ox)<sup>-2</sup> é quebrada, com a formação de FMN(ox)<sup>-2</sup> – '*flavin mononucleotide*' - e AMP<sup>-1</sup> – '*adenosine monophosphate*'. Neste caso, FAD(ox)<sup>-2</sup> não encontra uma correspondência direta do lado dos produtos e a sua  $\Delta G_f^o$  não é compensada. Ademais, tanto FAD como FMN foram definidos por R. Alberty como tendo  $\Delta G_f^o$  igual a zero, pelo que a aplicação direta desses valores, conjugados com a  $\Delta G_f^o$  de FAD(ox)<sup>-2</sup>, gera um conflito com as leis da conservação. Por este motivo, o conceito de unificar GC com RC é mais complexo do que pode inicialmente parecer, sob o risco de se basear em pontos de referência inconsistentes entre si.

Uma das possibilidades para evitar os conflitos entre estes pontos de referência inconsistentes reside na utilização exclusiva de GC ou RC, dependendo da reação em questão. Mais concretamente, RC seria aplicado a todas as reacções que pudessem ser completamente resolvidas por este método; caso uma ou mais espécies químicas constantes da reacção estivessem ausentes das tabelas de energias, então GC seria usado para o cálculo de  $\Delta G_r^o$  de toda a reação. Porém, esta forma de combinar ambos os métodos poderia também conduzir a violações das leis da termodinâmica, nomeadamente quando é feita a transição de reações isoladas para reconstruções metabólicas completas. Nestas circunstâncias, a existência de pontos de referência inconsistentes pode levar, por exemplo, à ocorrência de saldos energéticos diferentes de zero em ciclos que se esperaria que fossem fúteis. Considere-se o exemplo da Fig. 1.3.:



**Fig. 1.3.** – Exemplo de uma situação na qual as energias de Gibbs são estimadas através da aplicação conjunta de RC e GC, antes da introdução de correções no cálculo de  $\Delta G_r^\circ$  a partir de espécies químicas de referência. (Imagem adaptada de [20].)

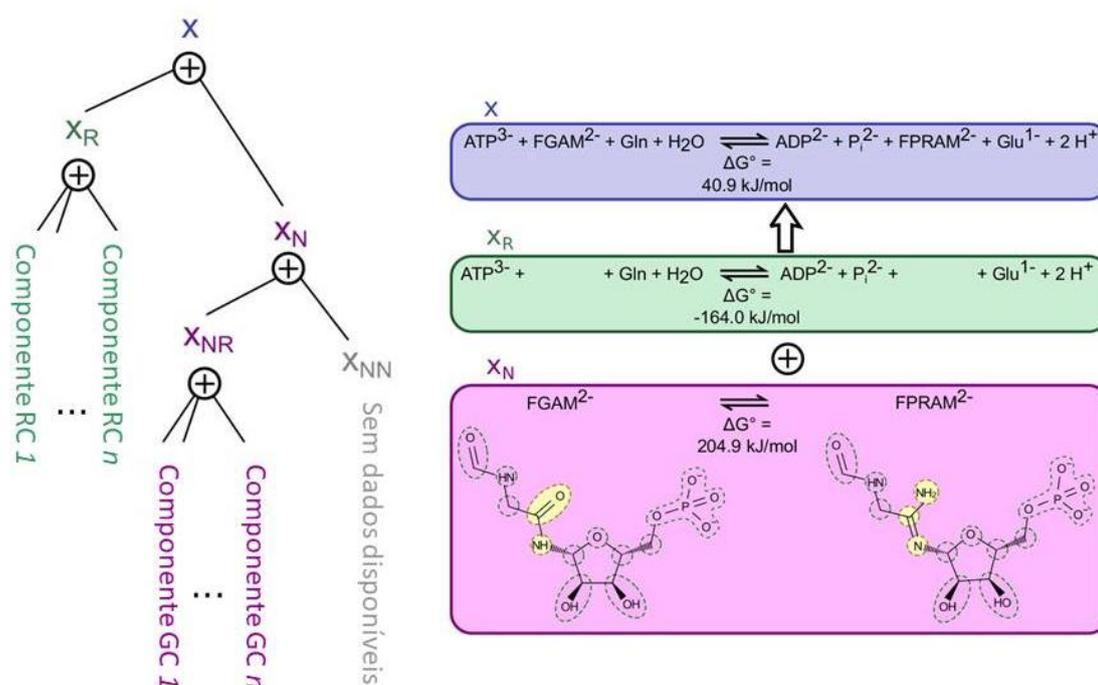
A Fig. 1.3. descreve um desses ciclos fúteis. As reações 1, 2, 3 e 4, quando combinadas, geram um ciclo fútil, isto é, o consumo dos reagentes e a formação dos produtos são sempre mutuamente compensadas. Num ciclo metabólico deste género, o saldo energético total deveria ser nulo. Seguindo primeiramente o método de RC, é possível estimar o valor de  $\Delta G_r^\circ$  para as reacções 1 e 2. Mas o metabolito 2-amino-3-oxobutanoato não consta das tabelas energéticas utilizadas em RC, pelo que as reacções 3 e 4 têm de ser resolvidas através de GC. O resultado final é que a soma de todos os  $\Delta G_r^\circ$  estimados perfaz um total de  $-6,0 \text{ kJ/mol}$ , constituindo uma violação à primeira lei da termodinâmica.

Em CC, este problema é resolvido através de um ajuste das energias livres de formação dos co-fatores não-elementares, de modo a que correspondam às suas respetivas estimativas através de GC. Assim, as tabelas energéticas foram corrigidas, de modo a que moléculas como  $\text{FAD}(\text{ox})^{-2}$  tenham sempre uma  $\Delta G_r^\circ$  correspondente ao seu valor obtido por GC. Consequentemente, todas as energias livres de formação calculadas relativamente a estes pontos de referência, foram também corrigidas, de forma a manter a coerência da tabela. Entre estas correções, o valor atribuído por GC a  $\Delta G_r^\circ$  de 2-amino-3-oxobutanoato é também retificado em conformidade. A principal desvantagem desta abordagem está no facto de o conjunto de pontos de referência ser fixo e limitado a um número reduzido de co-fatores, o que restringe a capacidade de cobertura deste método.

A extensão de reações abrangidas por *RC* poderia ser aumentada pela atribuição do estatuto de ponto de referência a outros compostos bioquímicos, mais invulgares, que habitualmente não constam das tabelas. Contudo, por questões práticas, torna-se difícil fazer uma listagem exaustiva de todos esses compostos.

Além das correções face às inconsistências descritas, em *CC* é introduzido um outro aspeto importante: apesar de *RC* ser prioritária face a *GC*, em vez de se usar exclusivamente uma ou outra para cada reação, todas as reações são divididas em duas partes – uma passível de ser resolvida por *RC*, e outra por *GC*. Esta divisão, na verdade, permite tratar cada reação como um vetor divisível em duas componentes, como se pode ver pela Fig. 1.4..

O vetor correspondente a uma reação,  $x$ , é decomposto em duas componentes – a primeira,  $x_R$  refere-se aos elementos cuja  $\Delta G^\circ$  faz parte das tabelas energéticas e pode ser resolvida por *RC*; a segunda,  $x_N$ , contém aqueles que terão de ser tratados por *GC*.  $x_N$  pode ainda ser subdividida, caso a reação contenha elementos que não possam ser tratados por *RC* nem *GC*. Note-se ainda que tanto  $x_R$  como  $x_N$  podem ser definidas como nulas.



**Fig. 1.4.** – Diagrama ilustrativo da decomposição do vetor de uma reação nas respetivas componentes *RC* e *GC*. É ainda visível um exemplo da decomposição de uma reação concreta, para a qual ambas as componentes são diferentes de zero. (Imagem adaptada de [20].)

## Introdução

Por tudo isto, *CC* é já um método robusto, capaz de produzir estimativas com boa precisão para um elevado número de reações e passível de ser transposto para a análise de vias metabólicas completas. Contudo, os próprios autores indicam que este método permanece dependente da qualidade dos dados experimentais presentes nas bases de dados utilizadas. De resto, são ainda sugeridas hipotéticas camadas intermédias entre *RC* e *GC* que poderiam melhorar a qualidade das estimativas produzidas.

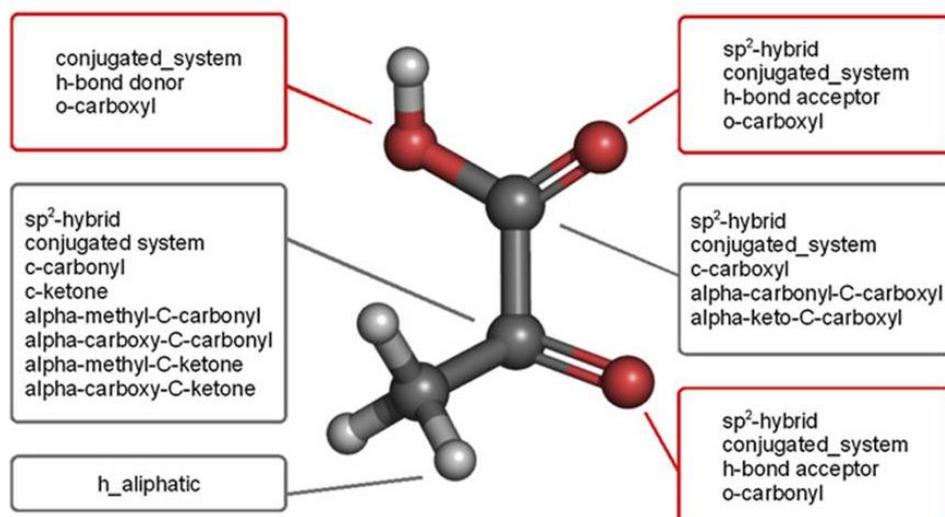
---

### 1.2.4. ITERS – ‘INFERRING GIBBS ENERGY CHANGES FROM REACTION SIMILARITIES’

---

*IGERS* é uma abordagem que diverge consideravelmente das apresentadas nas secções 1.2.1. a 1.2.3.. Baseia-se no princípio de que reações quimicamente similares deverão caracterizar-se por valores de  $\Delta G_r^o$  semelhantes. O ponto-chave deste método está na capacidade de quantificar as semelhanças entre duas reações. Pretende-se assim comparar uma reacção de  $\Delta G_r^o$  desconhecida com um conjunto predefinido de outras reações – para as quais se conhece  $\Delta G_r^o$  – e calcular um índice de semelhança para cada caso. No final do processo, é atribuído à reacção em estudo o valor de  $\Delta G_r^o$  daquela com a qual o índice de semelhança é mais elevado.<sup>[17]</sup>

*IGERS* baseia-se também na representação de cada reacção através de um vetor binário, mas neste caso com a sua caracterização química detalhada. A construção deste vetor requer a existência de uma matriz que faça o mapeamento correto da transição dos átomos dos reagentes para os átomos dos produtos. O método pressupõe ainda a definição de uma série de atributos físico-químicos, que proporcionam uma caracterização completa e consistente de cada metabolito presente numa reacção. A Fig. 1.5. exemplifica a atribuição destas características a uma molécula de piruvato.

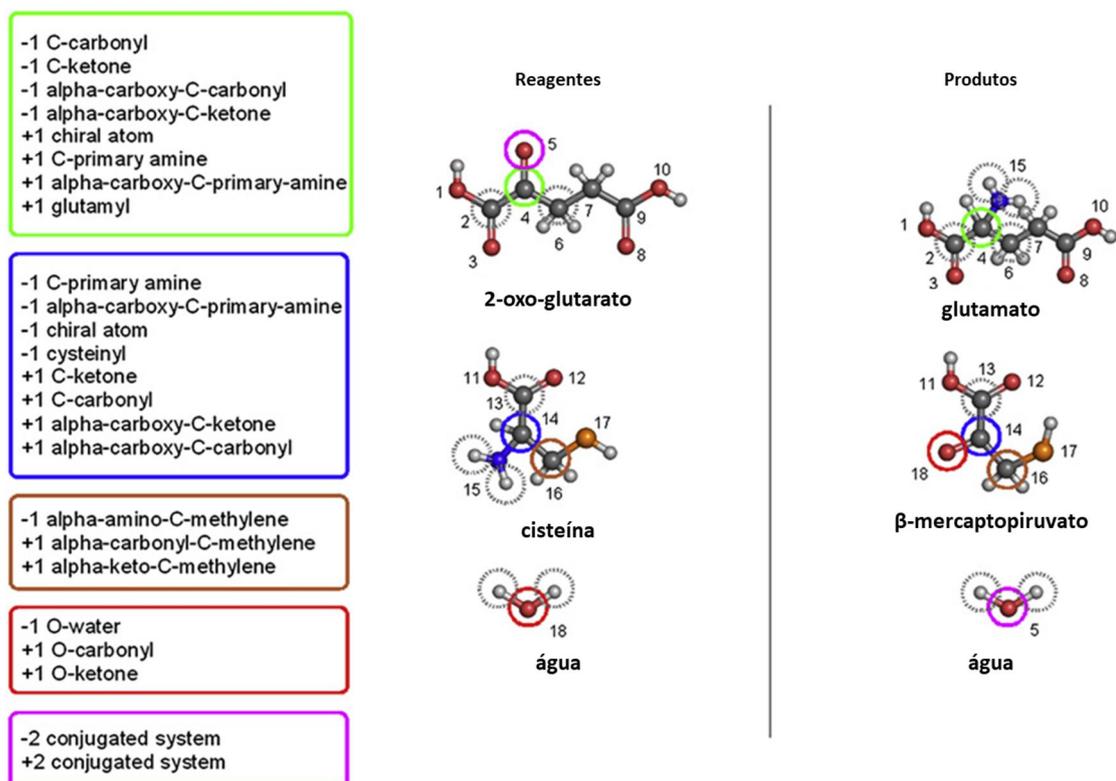


**Fig. 1.5.** – Enumeração dos atributos químicos que caracterizam uma molécula de piruvato. Estas características vão desde o tipo de ligação química que une cada átomo aos seus vizinhos, aos vários tipos de grupos funcionais presentes na molécula. No total, para o piruvato são definidos 38 atributos. (Os termos e expressões utilizados para definir os atributos químicos são os que foram originalmente definidos pelos autores de *IGERS*. Neste trabalho optou-se por não fazer qualquer tradução a este respeito, mantendo-se a nomenclatura original.) (Imagem adaptada de [17].)

O mecanismo de construção do vetor da reação é ilustrado na **Fig. 1.6.** Na transaminação da cisteína, o oxigênio na molécula de água vai participar na formação de uma cetona nos produtos, o  $\beta$ -mercapto-piruvato. Subsequentemente, forma-se uma nova molécula de água, a partir do oxigênio proveniente do 2-oxo-glutarato. Note-se que as mudanças nos atributos químicos derivadas destas transformações não são iguais e que é necessário que as matrizes de transição as traduzam corretamente. Ao todo, na ilustração da **Fig. 1.6.**, estão exemplificadas 24 das alterações aos atributos químicos entre reagentes e produtos, num total de 61. Todas essas alterações serão depois identificadas no vetor a construir.

Contudo, ao longo de todo o seu trabalho, os autores contabilizaram um total de 1274 tipos de alterações diferentes, traduzidas num vetor com 1274 elementos (1274 *bits*), cada um correspondente a uma das possibilidades. O valor de cada um desses elementos pode ser de **0** ou de **1**, dependendo de a correspondente alteração nos atributos ocorrer (**1**) ou não (**0**) naquela reação. No caso da reação ilustrada na **Fig. 1.6.**, os 61 *bits* do vetor referentes às alterações ocorridas teriam o valor **1**. Os restantes seriam preenchidos a **0**.

## Introdução

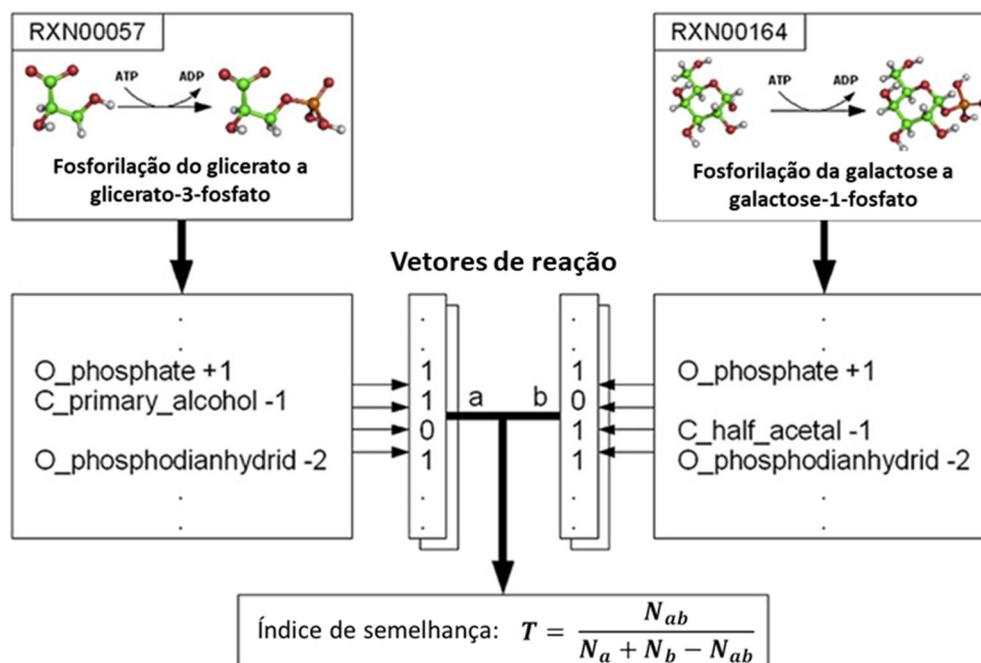


**Fig. 1.6.** – Representação esquemática da identificação de alterações nos atributos químicos entre reagentes e produtos numa transaminação da cisteína para o 2-oxo-glutarato. Estas alterações nos atributos são enumeradas nas caixas de contorno colorido à esquerda (“+” – atributo adquirido, “-” atributo perdido) e referem-se aos átomos rodeados com circunferências da cor correspondente. Os círculos a tracejado indicam átomos que sofrem pelo menos uma alteração de atributos, enquanto os átomos numerados são aqueles que não sofrem nenhuma alteração. Ao todo, esta transformação envolve 61 alterações de atributos químicos. (Os termos e expressões utilizados para definir os atributos químicos são os que foram originalmente definidos pelos autores de *IGERS*. Neste trabalho optou-se por não fazer qualquer tradução a este respeito, mantendo-se a nomenclatura original.) (Imagem adaptada de [17].)

Finalmente, o índice de semelhança entre reações é calculado através do nível de concordância entre os respetivos vetores. Para isso, o método eleito é o coeficiente de Tanimoto,  $T$ :

$$T = \frac{N_{ab}}{N_a + N_b - N_{ab}} \quad (\text{Equação 1.19.})$$

Neste coeficiente,  $N_a$  corresponde ao número de *bits* a **1** no vetor  $\mathbf{a}$ ,  $N_b$  o número de *bits* a **1** do vector  $\mathbf{b}$ , e  $N_{ab}$  o número de *bits* a **1** em comum entre o vetor  $\mathbf{a}$  e o vetor  $\mathbf{b}$ . (Note-se que os *bits* a **0** não são tidos em conta nesta operação.) Assim,  $T$  igual a um significa que os vetores são exactamente iguais;  $T$  igual a zero corresponde a vetores completamente diferentes. A Fig. 1.7. transmite um exemplo desta operação.



**Fig. 1.7.** – Exemplo do cálculo por *IGERS* do índice de semelhança entre duas reações através do coeficiente de Tanimoto. As caixas localizadas abaixo da ilustração de cada reação mencionam apenas 3 das alterações nos atributos químicos de cada reação. Tendo por base unicamente estas 3 alterações, teríamos:  $N_a = 3$ ,  $N_b = 3$ ,  $N_{ab} = 2$  e, conseqüentemente,  $T = 0,5$ . (Os termos e expressões utilizados para definir os atributos químicos são os que foram originalmente definidos pelos autores de *IGERS*. Neste trabalho optou-se por não fazer qualquer tradução a este respeito, mantendo-se a nomenclatura original.) (Imagem adaptada de [17].)

Outra nota importante no que diz respeito a *IGERS* prende-se com o sentido assumido para cada reação: o valor de  $T$  será diferente consoante uma dada reação,  $a$ , seja comparada com outra reação no seu sentido direto,  $b$ , ou com essa reação no sentido inverso,  $b'$ . Ou seja  $T(a,b) \neq T(a,b')$ . Da mesma forma, a inversão do sentido de uma reação, implica inverter as alterações nos atributos. Veja-se mais uma vez o exemplo dado pela ilustração da Fig. 1.6.: a alteração “+1 *O-ketone*” no sentido direto, seria convertida em “-1 *O-ketone*” no sentido inverso, sendo que estas são alterações codificadas em *bits* distintos no vetor binário. Daqui pode concluir-se a existência de pelo menos mais dois casos particulares com relevância na interpretação do coeficiente de Tanimoto:  $T(a,b) = T(a',b')$  e  $T(a,b') = T(a',b)$ .

### 1.2.5. OUTROS MÉTODOS COMPUTACIONAIS

Vários outros métodos têm já sido apontados como soluções para o cálculo da energia livre de Gibbs em reações bioquímicas. Entre os que serão abordados nas

## Introdução

próximas linhas, contam-se métodos com o dado comum de serem muito exigentes a nível computacional, apesar de produzirem resultados com elevada exatidão, obtidos com base em princípios físico-químicos rigorosos e bem conhecidos, que viabilizam uma cobertura de praticamente 100 % das reações metabólicas. Por este motivo, é também comum encontrar na literatura a sugestão de poderem ser aplicados enquanto métodos complementares, nomeadamente em casos para os quais não seja possível a obtenção de estimativas através dos métodos já descritos. Adicionalmente, as abordagens que se seguem podem também possibilitar a validação de valores experimentais obtidos a partir de um número reduzido (ou mesmo único) de medições.<sup>[20, 40]</sup>

### 1.2.5.1. MECÂNICA MOLECULAR E DINÂMICA MOLECULAR

---

Mecânica Molecular (*MM*) e Dinâmica Molecular (*MD*) são termos que se referem a métodos de simulação computacional que permitem estimar determinadas propriedades moleculares, normalmente relacionadas com aspetos estruturais e energéticos. Ambos se baseiam em funções da energia potencial deduzidas a partir da mecânica clássica, bem como em parâmetros inferidos pela via experimental (por exemplo, espetrometria de infravermelhos ou cristalografia de raios-X) ou pelo cálculo, de acordo com os princípios da mecânica quântica. É frequente a atribuição da expressão Campo de Forças ao conjunto das funções da energia potencial e dos parâmetros que lhe estão associados.

*MM* é normalmente aplicado quando se pretende calcular as energias relativas entre diferentes conformações geométricas para a mesma molécula. Estas diferentes conformações estão habitualmente relacionadas com fenómenos como rotações em torno de certas ligações químicas ou a formação de complexos intermoleculares. Frequentemente, *MM* é aplicado tendo em vista o cálculo de mínimos energéticos, pelo que também poderá ser utilizado em conjugação com algoritmos de otimização.<sup>[41]</sup>

As simulações computacionais por *MD* constituem uma metodologia que permite uma modelação detalhada de um sistema à escala atómica e são uma ferramenta poderosa, largamente aplicada em química e em física. Esta técnica permite acompanhar a evolução do sistema em estudo ao longo do tempo e fazer previsões sobre as propriedades estáticas ou dinâmicas das mais variadas substâncias a partir de uma monitorização das interações entre moléculas. Estas simulações operam através da resolução numérica das equações de movimento que descrevem os movimentos

executados pela molécula, fornecendo assim informação sobre propriedades estruturais e termodinâmicas. Através de *MD* é então possível coligir informação sobre a magnitude das variações de posição e de velocidade, bem como sobre a forma como estas variam ao longo do tempo. Dentro deste tipo de abordagens, é ainda possível referir o método de Monte Carlo, focado principalmente nas mudanças de posição, mas fornecendo pouca informação sobre a sua dependência face ao decorrer do tempo. Dependendo do nível de rigor desejado, as equações de movimento a serem resolvidas por estes métodos podem seguir o modelo clássico de Newton ou derivar para modelos mais complexos, tais como o movimento estocástico de Langevin, o modelo Browniano, ou mesmo combinações entre métodos de mecânica clássica e mecânica quântica.<sup>[42]</sup>

Estas metodologias revelam-se particularmente úteis no estudo de reações que decorrem em ambientes aquosos, como é o caso da grande maioria das reações metabólicas. Em particular, possibilitam a simulação pormenorizada das interações dos reagentes com o solvente e a obtenção de informação sobre aspetos como a eventual predominância de uma forma ionizada sobre outras, o que pode ser preponderante na análise de reações onde intervém grupos ionizáveis ou cuja termodinâmica seja especialmente influenciável por variações no pH, ainda que ligeiras. Contudo, as equações para a energia potencial utilizadas nestes métodos dificilmente podem ser adaptadas ao estudo de casos que envolvam a quebra e formação de novas ligações químicas.<sup>[43]</sup> Este último aspeto torna-se um obstáculo claro à aplicação destes modelos no contexto desta dissertação, pelo menos sem a sua conjugação com outras metodologias.

---

### 1.2.5.2. MÉTODOS QUÂNTICOS / *AB INITIO*

Os métodos quânticos ou *Ab Initio* debruçam-se sobre cálculos de mecânica quântica, derivados unicamente de princípios teóricos e ignorando qualquer tipo de conhecimentos resultantes da via experimental. Estes métodos procuram resolver a equação de Schrodinger, através da posição do núcleo dos átomos envolvidos e do respetivo número de eletrões – dos quais é possível obter densidades eletrónicas, energias e outras propriedades do sistema. Entre estes métodos, as abordagens de Hartree-Fock e Post-Hartree Fock são algumas das mais conhecidas e têm já sido referidas como parte integrante de soluções para a problemática das estimativas da

## Introdução

energia livre de Gibbs.

Os métodos *Ab Initio* oferecem a possibilidade de convergência para uma solução exata, embora o custo computacional de uma tal simulação seja frequentemente um entrave. Apesar do desenvolvimento de processadores e computadores modernos e com melhores capacidades, é habitual que o tempo requerido para computação pelos métodos *Ab Initio* leve a que em algumas aplicações específicas seja ainda preferível a opção por métodos menos rigorosos como *MM*.<sup>[44]</sup>

### 1.2.5.3. MÉTODOS HÍBRIDOS – MECÂNICA QUÂNTICA / MECÂNICA MOLECULAR

---

Os métodos computacionais baseados em mecânica quântica constituem a abordagem mais rigorosa e com resultados mais próximos da realidade quando se trata de estudar as propriedades de reações químicas. No entanto, no que toca a reações bioquímicas – que envolvem a participação de enzimas e, frequentemente, a transformação de proteínas e/ou ácidos nucleicos – o tamanho do sistema em análise facilmente ultrapassa os limites práticos da computação actual. Esta questão pode ser simplificada através do enfoque apenas no substrato e no local ativo da enzima, e ignorando o restante contexto. Ainda assim, o restante da estrutura proteica e o solvente podem alterar significativamente os parâmetros energéticos do sistema, pelo que se torna necessário encontrar uma forma de incluí-los na análise.

O aparecimento de métodos híbridos Mecânica Quântica – Mecânica Molecular (*QM-MM*) trouxe uma forma de responder ao problema. A base desta metodologia está na divisão do sistema entre uma região mais pequena – onde decorre a principal interação com o substrato – e que é tratada por *QM*; e o restante sistema, que é modelado pelos métodos menos exigentes da mecânica clássica. Estes métodos são também frequentemente designados como semi-empíricos.

Naturalmente que a principal vantagem destes métodos está em poderem ser aplicados a sistemas vastos a um custo computacional mais reduzido. Por outro lado, além do rigor reduzido face aos métodos puramente quânticos, os métodos semi-empíricos apresentam a desvantagem de regra-geral serem parametrizados para refletir as propriedades das moléculas no seu estado fundamental, o que torna difícil a sua aplicação em muitas reações químicas. A resposta a estas limitações está frequentemente na aplicação de métodos *Ab initio* ou de teoria dos funcionais da densidade na região que é resolvida por mecânica quântica. Ambas as abordagens

possibilitam um maior rigor quantitativo e a teoria dos funcionais da densidade fornecem uma forma mais prática de incluir correlações eletrônicas na análise.<sup>[45]</sup>

A aplicação de abordagens híbridas em múltiplos contextos está cada vez mais enraizada e são disso exemplo vários estudos em física e ciências de materiais<sup>[46]</sup>, bem como a sua utilização na descrição do comportamento de moléculas e reacções bioquímicas<sup>[43, 47, 50]</sup>, incluindo ao nível dos seus aspectos termodinâmicos<sup>[40, 48, 49]</sup>.

### 1.3. ESCOLHA DOS MÉTODOS A ANALISAR

---

Como evidenciado na secção 1.2., a variedade de métodos já desenvolvidos para estimar a variação da energia livre de Gibbs em reacções metabólicas é extensa e compreende métodos com elevada robustez científica e uma verdadeira aplicabilidade prática. Em particular, *CC* e *IGERS* demonstram não só ser versáteis em termos de capacidade de gerar estimativas válidas para novas reacções (isto é, com uma elevada abrangência), como ainda parecem ser capazes de o fazer com um grau de precisão deveras interessante (principalmente quando comparados diretamente com *GC*).

Independentemente dos méritos associados a cada um dos métodos previamente descritos, continua a haver espaço para melhorias. Enquanto métodos como os referidos na secção 1.2.5. produzem resultados extremamente precisos, a sua exigência computacional e a impossibilidade serem aplicados em tempo útil a um elevado número de reacções colocam-nos fora do âmbito de trabalhos voltados para a simulação e seleção de vias metabólicas completas. Por outro lado, os métodos estimativos já apresentados – *GC*, *CC*, *IGERS* – mostram-se capazes de produzir resultados interessantes e em tempo útil para um largo espectro de reacções, incluindo aquelas para cujos intervenientes as energias de formação são desconhecidas, e especialmente no que diz respeito ao grau de precisão com que o conseguem fazer, num curto espaço de tempo.

Não obstante, *GC* – e outros métodos que a este recorram, ainda que apenas parcialmente -, apresenta sempre a limitação de assumir sempre a aditividade pura das energias de formação dos vários grupos constituintes de cada molécula, o que frequentemente conduz a margens de erro demasiado grandes. Relativamente a *GC*, o

## **Introdução**

desenvolvimento de *CC* e *IGERS* permitiu acima de tudo um considerável estreitar das margens de erro. Apesar disso, note-se que mesmo em *CC*, qualquer reação que faça uso de *GC* – por inteiro ou de forma parcial – apresentará sempre uma margem de erro elevada, uma vez que estará sempre presente o erro associado à aditividade das energias de cada grupo. Além disso, a aplicabilidade do método continua a ser limitada pelas tabelas de energias de formação dos grupos constituintes, sendo nula quando aparecem reagentes ou produtos que incluam grupos que não estejam definidos à partida.

*IGERS* destaca-se por não depender – pelo menos de uma forma direta – das energias de formação dos compostos intervenientes nas reacções, nem da definição dos grupos funcionais nos quais podem ser decompostos. Uma vez que *IGERS* faz uma comparação direta das similaridades entre reacções, o resultado final depende unicamente do conhecimento acumulado sobre valores de  $\Delta G_r^0$  determinados experimentalmente (e, especialmente, do conhecimento disponível sobre reacções suficientemente semelhantes às que estejam em análise; tal disponibilidade influencia sempre o valor de *T* com o qual é possível trabalhar e produzir resultados com boa precisão).

É também interessante constatar algumas notas deixadas pelos autores de *CC* e de *IGERS*, referentes ao melhoramento dos trabalhos desenvolvidos e à possibilidade de complementação através de outros métodos. No caso de *IGERS*, os autores defendem que a sua abordagem poderia beneficiar de uma conjugação com *GC*. Sabendo da dificuldade de obter, a um ritmo elevado, novos dados experimentais sobre  $\Delta G_r^0$  de reacções bioquímicas que permitam alargar a aplicabilidade do método, a sua articulação com *GC* alargaria consideravelmente a abrangência proporcionada. Por outro lado, os autores de *CC* indicam *IGERS* como um possível nível intermédio a estabelecer entre *RC* e *GC*. Neste caso, presume-se que o algoritmo começaria por resolver através de *RC* todos os casos possíveis. Para os restantes casos, *IGERS* seria sempre prioritário em relação a *GC*: não sendo possível a resolução por *RC*, as reacções em análise seriam comparadas com uma ‘pool’ de reacções previamente estabelecida (muito possivelmente, as mesmas que funcionassem como a base de trabalho de *RC*) e para cada caso seria atribuído o valor de  $\Delta G_r^0$  da reação mais semelhante; só em último caso seria aplicado *GC* (por exemplo, no caso de não ser possível encontrar nenhuma reação com *T* igual ou superior a um valor que gere um resultado com uma margem de erro inferior à de *GC*).

Assim, definiu-se como alvo deste trabalho a exploração de uma hipotética relação de complementaridade entre *CC* e *IGERS*, ou outros métodos que, no seu conjunto, permitam combinar os conhecimentos obtidos pela via experimental com GC e com a comparação assente em similaridades químicas. Com isto pretende-se principalmente melhorar a precisão das estimativas produzidas, e, conseqüentemente, melhorar a perspectiva de aplicabilidade a vias metabólicas completas. Essencialmente, espera-se que as estimativas geradas apresentem um menor nível de dependência em relação ao princípio da aditividade das energias de formação dos grupos resultantes da decomposição de cada molécula.

### 1.4. DISPONIBILIDADE DE *IGERS* E MÉTODOS ALTERNATIVOS

---

O método *IGERS* está inteiramente disponível para consulta e *download* através da página <https://bitbucket.org/krother/deltaglib/overview>, onde é possível aceder a todo o código em Python 2.5, bem como a todos os respetivos ficheiros de instruções de utilização. Porém, o facto de a sua publicação ser já de 2010 leva a que haja algum desfasamento entre os recursos originalmente utilizados no seu desenvolvimento e os que estão atualmente disponíveis. Em concreto, a atribuição das energias livres de Gibbs após o cálculo dos índices de semelhança era feita com base nos dados disponíveis para um conjunto de 173 reações constantes da base de dados BioPath®. Esta base de dados é, pelo menos nos dias que correm, de acesso restrito, e está disponível através do endereço <https://www.mn-am.com/databases/biopath>, propriedade conjunta das empresas Altamira, LLC (Estados Unidos da América) e Molecular Networks GmbH (Alemanha). Apesar disto, o conjunto de reações, e respetivos dados experimentais, utilizado pelos autores de *IGERS* é disponibilizado numa lista anexa a [17].

Em todo o caso, é necessário notar que a versão original de *IGERS* está otimizada para consulta direta da base de dados BioPath®. Conseqüentemente, todo o processo de consulta dos dados necessários para computação decorreria de acordo com o formato sob o qual a informação estaria organizada naquela base de dados à data da publicação do trabalho. Esta consulta incluiria o acesso a ficheiros MDL com a estrutura de

## **Introdução**

moléculas e/ou reações (vulgarmente designados como ficheiros “.mol”; constituem um formato amplamente estabelecido para modelação informática de moléculas e reações químicas), bem como a sua posterior utilização como referência para a definição dos atributos químicos considerados pelo modelo.

Numa primeira fase, considerou-se a viabilidade de adaptar *IGERS* para a consulta de outra(s) base(s) de dados. (Nomeadamente, a mesma utilizada por *CC* – “*NIST-TECRDB – Thermodynamics of enzyme-catalyzed reactions database*”, pois, para efeitos da comparação que se pretende aqui levar a cabo, faz todo o sentido que ambos os métodos sejam alicerçados sobre numa base de dados comum, o que não se verifica nas aplicações isoladas dos mesmos que estão disponíveis na literatura.) No entanto, acontece que sem conhecimentos concretos da organização conceptual de *BioPath*® torna-se difícil e particularmente moroso compreender que informações estão a ser consultadas e de que modo são depois utilizadas pelo método. Além disso, seria necessária uma manipulação de porções extensas do código e em várias fases dos algoritmos de enumeração dos atributos químicos e de comparação. Todo este processo implicaria o risco acrescido de introduzir erros e inconsistências no método, com evidente prejuízo da validade dos resultados obtidos.

Nestas circunstâncias, procuraram-se alternativas que permitissem a comparação direta entre reações químicas e o estabelecimento de um índice de similaridade. À semelhança de *IGERS*, esse índice seria depois usado para ordenação num *ranking* de semelhanças químicas e atribuição de um valor concreto de  $\Delta G_r^o$ .

A *ChemAxon* é uma empresa dedicada à concepção de software, nomeadamente ao nível de soluções para utilizador final na investigação em química e biologia. Entre outras ferramentas desenvolvidas por esta empresa, o *JChem*® foi especialmente importante nesta dissertação, proporcionando um método alternativo eficaz e simples de utilizar, tendo em vista a comparação direta entre reações químicas e o cálculo de índices de semelhança.

A *Similarity Search* é uma das funcionalidades que integram o *JChem*® e visa a procura de estruturas químicas semelhantes entre si (ou o mais próximas possível), com base num índice de semelhança estabelecido pelo utilizador. Esta funcionalidade pode ser aplicada tanto a moléculas isoladas como a reações químicas e envolve a seleção do objeto da comparação (*‘query’*), bem como do universo de referências com o qual esse

objeto será comparado (*'targets'*). O índice de semelhança resultante de cada comparação é também obtido por um coeficiente de *Tanimoto* (embora após um processo mais complexo do que em *IGERS*, que inclui a comparação objetiva entre as estruturas de reagentes e produtos, e ainda a comparação entre os átomos que efetivamente sofrem transformações) e o conjunto final de todos os coeficientes obtidos após uma série de comparações pode ser seriado num *ranking*.

No final deste processo, e aplicando o mesmo princípio definido pelos autores de *IGERS*, o *target* correspondente ao índice de semelhança mais elevado (mais próximo de 1) será considerado como o mais parecido com a reação de *query* e as suas energias livres de Gibbs serão equiparadas. Uma *Similarity Search* pode ainda ser invocada a partir da linha de comandos do sistema operativo em uso, o que facilita a sua inclusão numa pequena aplicação a ser desenvolvida de raiz para este efeito. Esta aplicação envolve ainda a criação de ficheiros RXN (*“.rxn”*), referentes a todas as reações envolvidas neste processo. Numa comparação muito simplificada, estes ficheiros são o equivalente aos ficheiros MDL no que respeita à modelação informática de reações químicas. Descrevem tanto o bloco dos reagentes como o bloco dos produtos e condensam toda a informação relevante acerca da transformação dos primeiros nos segundos. (Outras informações detalhadas sobre ambos os formatos podem ser consultadas através do endereço: <https://docs.chemaxon.com/display/docs/MDL+MOLfiles%2C+RGfiles%2C+SDfiles%2C+Rxnfiles%2C+RDfiles+formats>.) Estes ficheiros funcionam como autênticas descrições pormenorizadas de cada reação e é através deles que a *Similarity Search* é conduzida, pelo que é necessário gerar ficheiros RXN tanto para todas as *queries* como para todos os *targets*.

Até à presente data – Outubro de 2018 - toda a informação teórica sobre estes métodos e sobre a forma como podem ser aplicados está disponível para consulta em <https://docs.chemaxon.com/display/docs/Similarity+search>. Em todo o caso, alguns destes pormenores serão mais detalhados ao longo dos próximos capítulos. Note-se ainda que, no âmbito desta dissertação, a utilização das ferramentas de software disponibilizadas pela *ChemAxon* foi feita com recurso a uma licença para utilização em contexto académico, de acordo com os termos estabelecidos pela empresa (<https://chemaxon.com/academic-license>).

### 1.5. OBJETIVOS

---

Finalizada a contextualização deste trabalho, torna-se necessário enumerar de forma sucinta e objetiva quais as tarefas que lhe estão implícitas, bem como elucidar o conjunto de objetivos traçados. Tal como explicado na secção 1.3., pretende-se essencialmente utilizar *CC* e um método de princípios semelhantes aos de *IGERS* para estimar a energia livre de Gibbs para um mesmo conjunto de reacções metabólicas. Os resultados de ambos os métodos serão posteriormente alvo de uma avaliação comparativa, procurando perceber até que ponto o segundo poderá colmatar as fragilidades do primeiro.

Assim, os próximos capítulos pautar-se-ão de acordo com os propósitos específicos que se seguem:

#### CAPÍTULO 2 - DESCRIÇÃO DAS METODOLOGIAS ADOPTADAS E DAS VIAS METABÓLICAS ESTUDADAS

**Objetivos:** Desenvolver uma aplicação em Python 2.7 para gerir a utilização das funcionalidades do *JChem* e construir um método baseado em semelhanças químicas; testar o funcionamento de ambos os métodos para uma amostra de 200 reacções isoladas e para 3 vias metabólicas completas.

#### CAPÍTULO 3 - RESULTADOS E DISCUSSÃO

**Objetivos:** Apresentar os resultados obtidos e avaliar comparativamente as vantagens e limitações evidenciados por cada método; justificar os resultados obtidos e enunciar possíveis formas de melhoramento dos métodos aplicados.

#### CAPÍTULO 4 - CONCLUSÕES E TRABALHO FUTURO

**Objetivos:** Estabelecer um balanço final dos resultados e ilações tiradas da aplicação dos dois métodos; propor termos concretos para o estabelecimento de uma verdadeira relação de complementaridade entre as metodologias abordadas ou para futuros trabalhos que visem o seu aperfeiçoamento.

## **2. DESCRIÇÃO DAS METODOLOGIAS ADOTADAS E DAS VIAS METABÓLICAS ESTUDADAS**

---



A análise e comparação dos resultados produzidos pelos métodos usados neste trabalho implica uma explicação mais detalhada acerca de cada um deles, bem como dos recursos sobre os quais se baseiam. É por isso importante fazer algumas referências mais pormenorizadas sobre diversos aspetos, que incluem a base de dados NIST-TECRDB, a *'pipeline'* que gere a criação de ficheiros RXN e a sua comparação sistemática, e ainda a escolha dos conjuntos de reações sobre os quais os métodos foram aplicados.

Quanto à base de dados NIST-TECRDB, é necessário explicar de que forma é utilizada por cada um dos métodos e como é que isso poderá influenciar os resultados obtidos. Relativamente aos processos seguidos por *CC* e pelo método desenvolvido neste trabalho – que passaremos a designar por *ReactionFingerprinter* – será concedida uma maior atenção a este último, uma vez que *CC* é já um método amplamente explorado e devidamente detalhado pelos seus autores<sup>[18, 20]</sup>. Será ainda importante especificar o processo de escolha das reações e vias metabólicas que serviram de amostra para comparação dos métodos e quais os critérios tidos em conta para a sua seleção.

## 2.1. A BASE DE DADOS NIST-TECRDB

---

A base de dados NIST-TECRDB (<https://randr.nist.gov/enzyme/>) – daqui em diante abreviada para *TECR* – consiste num vasto conjunto de informações sobre os parâmetros termodinâmicos de reações bioquímicas catalisadas por enzimas. Estas informações são essencialmente constantes de equilíbrio aparentes e entalpias molares (determinadas por calorimetria) e resultam da análise termodinâmica experimental dessas reações bioquímicas. Estes dados provêm de cerca de 1000 publicações, relacionadas com o estudo de aproximadamente 400 reações enzimáticas.<sup>[51]</sup> Daqui é importante tirar desde já uma nota importante: para muitas reações (quase todas, na verdade) existem dados provenientes de diversos trabalhos, frequentemente baseados em abordagens diferentes entre si, e que resultam em discrepâncias, maiores ou menores, entre as constantes de equilíbrio aparentes obtidas. Assim, fica desde já implícita a necessidade de aplicar alguma forma de filtrar ou normalizar os parâmetros termodinâmicos disponibilizados nesta base de dados.

## ***Descrição das metodologias adoptadas e descrição das vias metabólicas estudadas***

A base de dados TECR contém ainda várias páginas de discussão acerca dos fundamentos da termodinâmica de reacções catalisadas por enzimas, de onde constam, por exemplo, as definições das principais quantidades termodinâmicas de interesse. É o caso da constante de equilíbrio aparente,  $K'$ , e da entalpia molar de reacção obtida por calorimetria,  $\Delta H_r$  (cal), das quais é possível obter a variação molar da energia livre de Gibbs padrão,  $\Delta G_r^0$ . São ainda debatidas as diferenças entre a definição de reagentes e espécies intervenientes nas reacções bioquímicas. Neste sentido, as reacções bioquímicas são descritas em termos do somatório das espécies químicas envolvidas (por exemplo,  $\text{ATP}^{4-}$ ,  $\text{HATP}^{3-}$  ou  $\text{MgATP}^{2-}$ ), por sua vez referidos simplesmente como reagentes (por exemplo, ATP). Distinções deste género são mantidas em toda a base de dados e são importantes na manutenção da consistência entre os dados relativos às diferentes reacções. A base de dados TECR procura ainda manter a definição precisa das quantidades físico-químicas e a consistência na sua aplicação, bem como a padronização das nomenclaturas segundo as recomendações da IUPAC e da IUBMB. Também a ordenação das reacções ao longo da base de dados é feita por ordem do número atribuído pela *Enzyme Commission* a cada reacção. A base de dados permite ainda ao seu utilizador efetuar pesquisas com base nos reagentes envolvidos, no nome e número da enzima catalisadora, ou mesmo da solução tampão,  $pH$ , co-fatores e métodos utilizados no estudo.<sup>[51]</sup>

Em suma, a base de dados TECR apresenta-se como um conjunto robusto e consistente de dados termodinâmicos sobre reacções metabólicas, perfeitamente passível de ser utilizado como suporte teórico para métodos inovadores para estimar energias livres de Gibbs. Contudo, existem ainda algumas particularidades relativamente à sua aplicação nos métodos aqui abordados e que serão detalhadas nas próximas secções.

---

### 2.1.1. BASE DE DADOS TECR EM *COMPONENT CONTRIBUTION* – LINEARIZAÇÃO E CONJUGAÇÃO COM INFORMAÇÕES ADICIONAIS

---

$CC$  integra  $RC$  e  $GC$  numa plataforma única, através de uma técnica de regressão linear aplicada por camadas. Esta abordagem possibilita a utilização de dados experimentais precisos, posteriormente complementados com *group contributions* quando aqueles são inexistentes. Para tal, faz uso de uma matriz estequiométrica,  $S \in \mathbb{R}^m$

$x^n$ , que descreve as reações previamente analisadas pela via experimental, bem como a correspondente lista de energias livres de Gibbs observadas,  $\Delta G_{r,obs}^0 \in \mathbb{R}^n$ .<sup>[20]</sup> Todos estes dados termodinâmicos provêm da base de dados TECR e de tabelas de energias livres de formação de diversos compostos publicadas por Thauer *et al* e Alberty *et al*<sup>[20, 37, 39]</sup>. Contudo, em TECR são comuns as descrições de reações analisadas por diferentes abordagens experimentais, cujas condições podem variar, como é o caso do pH, da temperatura, ou da força iônica. Daqui decorre a existência de situações como a ocorrência de diferentes resultados para a mesma reação, gerando a necessidade de aplicar um passo de pré-processamento. Em termos gerais, este pré-processamento visa a normalização das constantes de equilíbrio disponíveis em TECR, bem como de algumas energias livres de formação.

O modelo de regressão utilizado em *RC* baseia-se na primeira lei da termodinâmica: a variação total na energia livre de Gibbs padrão de uma reação que ocorra em mais do que um passo corresponde ao somatório das energias livres de Gibbs de cada um desses passos, desde que todos ocorram sob as mesmas condições. Então, se  $\Delta G_f^0 \in \mathbb{R}^m$  for o vetor das energias livres de formação padrão dos compostos de *S*, as energias de Gibbs padrão das reações de *S* são dadas pela *Equação 2.1.*:

$$\Delta G_r^0 = S^T \cdot \Delta G_f^0 \quad (\text{Equação 2.1.})$$

Na prática, nem sempre esta abordagem traduz a realidade, uma vez que os valores de  $\Delta G_{r,obs}^0$  em TECR são obtidos por métodos que estão sempre sujeitos a ruído e às limitações instrumentais, sendo ainda que a força iônica do meio, *I*, nem sempre é conhecida.

A energia livre de Gibbs padrão,  $\Delta G_{r,x} \in \mathbb{R}$ , de reações para as quais se desconhecem resultados experimentais, com vetor estequiométrico  $x \in \mathbb{R}^m$  pode ser estimada por *RC* através de:

$$\Delta G_{r,rc,x}^0 = x^T \cdot \Delta G_f^0 \quad (\text{Equação 2.2.})$$

Geralmente, a primeira lei da termodinâmica implica que a energia de Gibbs padrão de uma combinação linear de reações é equivalente a uma combinação linear aplicada às respectivas energias livres de cada reação. Se  $x = Sw$ , então  $\Delta G_{r,rc,x}^0 = w^T \cdot \Delta G_{r,obs}^0$ .

## ***Descrição das metodologias adoptadas e descrição das vias metabólicas estudadas***

A abordagem de *RC* pode ser usada para estimar energias livres de Gibbs para qualquer reação com  $x$  no domínio de  $S$ ,  $R(S)$ , ou seja, reações que sejam combinações lineares de reações que efetivamente façam parte de  $S$  (e que, portanto, possuam pelo menos uma solução para  $x = Sw$ ). Qualquer componente de  $x$  que não faça parte de  $R(S)$  fica forçosamente excluído do âmbito de *RC*.  $S$  representa, apesar de tudo, uma pequena fracção de todo o universo de reações metabólicas e a maioria destas não pode ainda ser analisada por *RC*. Por exemplo, a reação catalisada pela enzima CMP fosfohidrolase –  $\text{CMP} + \text{H}_2\text{O} \rightleftharpoons \text{citidina} + \text{P}_i$  – não faz parte do conjunto de reações analisadas em *TECR*. Embora tanto a CMP como a citidina apareçam noutras reações (nomeadamente:  $\text{CMP} + \text{ATP} \rightleftharpoons \text{CDP} + \text{ADP}$ ;  $\text{citidina} + \text{H}_2\text{O} \rightleftharpoons \text{uridina} + \text{NH}_4$ ), não existe nenhuma combinação linear de reações que permita estimar  $\Delta G_r^0$  para a reação em questão.

Se o foco de *RC* está na utilização de energias livres de reações e de formação já conhecidas (seja de forma direta, seja por combinações lineares), em *GC* a abordagem não deixa de ser semelhante, embora direcionada para as energias livres de formação de grupos de átomos específicos. Essencialmente, cada composto metabólico representado em  $S$  é decomposto com base num conjunto predefinido de subgrupos estruturais. Cada uma destas decomposições é detalhada numa matriz de grupos,  $G \in \mathbb{R}^{m \times g}$ , e  $\Delta G_r^0$  é equiparada a uma combinação linear de contribuições das energias livres padrão,  $\Delta G_g^0$ , dos grupos representados em  $G$ . No final, o modelo de regressão linear para *GC* é traduzido por:

$$\Delta G_r^0 = S^T G \cdot \Delta G_g^0 \quad (\text{Equação 2.3.})$$

$S^T G \in \mathbb{R}^{n \times g}$  descreve as decomposições em grupos das reações em  $S$ , isto é, a estequiometria dos grupos consumidos e produzidos em cada reação. A estimativa por *GC* para  $x$  pode ser explicada como:

$$\Delta G_{r_{gc},x}^0 = x^T G \cdot \Delta G_{g_{gc}}^0 \quad (\text{Equação 2.4.})$$

Embora mesmo por *GC* não seja ainda possível decompor todos os compostos que intervêm em reações bioquímicas, a sua cobertura é, evidentemente, muito maior do que em *RC*. Contudo, assumir a aditividade pura das energias livres de formação dos diferentes grupos nem sempre conduz a resultados com a correção desejada - os autores de *CC* estimam o erro quadrático médio de *GC* em cerca de 6,8 *kJ/mol*. Também à

semelhança de  $RC$ , a cobertura de  $GC$  está limitada às possíveis combinações lineares das decomposições das reações analisadas experimentalmente, ou seja,  $G^T x \in R(G^T S)$ .

No que diz respeito à combinação de  $RC$  com  $GC$ , qualquer vetor  $x$  em  $\mathbb{R}^m$  pode ser decomposto num componente  $x_R$  no domínio de  $S$  e num componente  $x_N$ , ortogonal ao primeiro, no espaço nulo de  $S^T$  (ver Fig. 1.4.). (Neste contexto, o espaço nulo de  $S^T$  representa todo o conjunto de espécies químicas cuja  $\Delta G_f^0$  não pode ser obtida através de  $S^T$ .)  $CC$  aplica o método mais preciso de  $RC$  para resolver  $x_R$  e recorre a  $GC$  apenas para resolver  $x_N$ . A estimativa para a energia livre de Gibbs de  $x$  é obtida pelo somatório das contribuições de  $RC$  e  $GC$ :

$$\Delta G_{r,cc,x}^0 = x^T_R \cdot \Delta G_{f,rc}^0 + x^T_N G \cdot \Delta G_{g,gc}^0 \quad (\text{Equação 2.5.})$$

O componente  $x_N$  no espaço nulo de  $S^T$  só pode ser resolvido no caso de  $G^T x_N$  fazer parte do domínio de  $G^T S$ , ou seja, no espaço coberto pelas contribuições das decomposições em grupos. É por isso fundamental que  $x_N = x_{NR}$ , onde  $x_{NR}$  corresponde ao componente  $x_N$  de  $R(G^T S)$ . Se  $x_N$  possuir um componente  $x_{NN} = x_N - x_{NR} \in N(S^T G)$  diferente de zero, a reação geral não poderá ser analisada de todo por  $CC$ . (Um exemplo simples em que  $x_{NN} \neq 0$  é o de uma reação onde ocorra a formação de um grupo invulgar. Se esse grupo não constar do conjunto de reações de  $TECR$  - ou se for sempre conservado em todas as reações em que aparece -, então a sua contribuição é desconhecida.)  $CC$  baseia-se assim em duas camadas de projeções ortogonais: a primeira em que  $x$  é decomposto em  $x_R$  e  $x_N$ ; e a segunda em que  $x_N$  é decomposto em  $x_{NR}$  e  $x_{NN}$  (ver Fig. 1.4.).

Em todo o caso, note-se que há particularidades matemáticas da aplicação deste método que foram aqui abreviadas. A formulação teórica de  $CC$  foi devidamente explicitada pelos seus autores previamente e não há necessidade fazer uma descrição exhaustiva de aspetos que ultrapassam o entendimento necessário à sua aplicação e análise dos respetivos resultados.

---

2.1.2. UTILIZAÇÃO DE TECR EM *REACTION FINGERPRINTER*

---

De acordo com o que foi debatido nas secções 1.4. e 1.5., optou-se por utilizar a base de dados TECR também para a aplicação do método de estimativa aqui desenvolvido – *ReactionFingerprinter*. Para isso utilizou-se uma função integrante de *CC* e que permite a criação de uma cópia dos dados de todas as reações e energias de formação que constituem a base sobre a qual *CC* foi desenvolvido.

Esta função, *savecsv*, faz parte da Classe *TrainingData*, inserida no módulo *training\_data.py*. O seu propósito reside na consulta de todos os dados relativos às reações de TECR (e algumas energias livres de formação adicionais inseridas pelos autores) - já após o seu pré-processamento e normalização - e armazenamento num ficheiro “.csv”. Deste ficheiro constam a equação química de cada reação (com cada composto a ser codificado pela identificação atribuída pela base de dados KEGG: *Kyoto Encyclopedia of Genes and Genomes*<sup>[52]</sup>), as condições sob as quais cada ensaio foi conduzido (sempre que disponíveis: temperatura, pH e força iónica), a respetiva referência atribuída pela TECR ou pelos autores de *CC*, a variação da energia livre de Gibbs determinada sob as condições do ensaio (em *kJ/mol*) e a variação da energia livre de Gibbs para condições padrão (em *kJ/mol*).

O facto de o cerne de *ReactionFingerprinter* ser a comparação direta entre as características de cada reação (e não de medições de qualquer tipo de parâmetro termodinâmico) elimina a necessidade de contar com os dados da totalidade das reações envolvidas em *CC*, onde são frequentes várias repetições da mesma reação. Por este motivo, nesta fase foi feita uma seleção das reações disponíveis (secção 2.2.1.), escolhendo-se apenas uma única para cada caso, tendo apenas o simples cuidado de não repetir equações químicas semelhantes entre si. Este passo evita a criação de ficheiros RXN contendo informações repetidas e, acima de tudo, previne que a jusante, e para um grande número de casos, sejam repetidamente realizadas comparações entre as mesmas reações. Estas comparações fúteis resultariam num maior tempo de comparação e no armazenamento de informação repetida que teria de ser filtrada de alguma forma. Não obstante, conservaram-se à parte os dados respeitantes a todas as reações e energias de formação utilizadas em *CC*, uma vez que as diferenças existentes entre estas podem influenciar a atribuição final de uma estimativa para a variação da energia livre de Gibbs.

## 2.2. REACTION FINGERPRINTER – JCHEM®, ‘REACTION FINGERPRINTS’ E ‘SIMILARITY SEARCH’

---

O desenvolvimento do método *ReactionFingerprinter* teve por objetivo mimetizar o mesmo tipo de abordagem proposta pelos autores de *IGERS*<sup>[17]</sup>. Resumidamente, reações semelhantes deverão caracterizar-se por variações semelhantes da energia livre de Gibbs. Ademais, sendo a variação da energia livre de Gibbs uma função de estado – o seu resultado depende exclusivamente dos estados inicial e final do sistema –, a busca de semelhanças entre reações pode efetivamente produzir um método capaz de produzir estimativas próximas da realidade.

Esta ideia está inclusive na génese de *GC*, onde se teoriza que o ponto central da variação da energia livre de Gibbs está na contabilidade das energias de formação dos grupos de átomos consumidos e criados – aqueles que se mantêm de ambos os lados da equação acabam por anular-se mutuamente. Neste sentido, um mesmo mecanismo de reação, ainda que possa coincidir em reações com intervenientes muito diferentes, deverá conduzir a uma variação semelhante na energia livre de Gibbs. Este princípio poderá ser especialmente válido no caso de reações que, mesmo que catalisadas por enzimas diferentes, decorrem com intervenção de algumas moléculas ou co-fatores mais comuns no metabolismo, como sejam  $\text{NAD}^+/\text{NADH}$ ,  $\text{ATP}/\text{ADP}/\text{AMP}$  ou a coenzima A. O mecanismo de reação depende frequentemente do aporte energético motivado por reações redox ou fosforilação/desfosforilação para que sejam levadas a cabo transformações relativamente comuns noutras moléculas. Contudo, nem sempre existem combinações lineares que permitam a sua análise por *RC*, ou então envolvem o consumo ou criação de grupos que não estão previstos em *GC*. Esta lógica poderá ainda ser aplicável, ainda que em menor grau, a outras reações que não dependam da intervenção daqueles co-fatores.

As dificuldades encontradas na implementação directa de *IGERS* (ver secção 1.4.) levaram à necessidade de encontrar uma alternativa apropriada e que assentasse igualmente na pesquisa por similaridades químicas entre reações. Essa alternativa poderia então ser aplicada nesta dissertação segundo princípios teóricos semelhantes aos de *IGERS*: deveria garantir que o foco da abordagem continuasse a ser a comparação das reações a analisar com um conjunto de reacções para as quais a energia livre de

## ***Descrição das metodologias adoptadas e descrição das vias metabólicas estudadas***

Gibbs fosse já conhecida pela via experimental. É aqui que entra em cena a utilização de *JChem*®, uma ferramenta informática que proporciona ao seu utilizador a possibilidade de realizar a procura de estruturas com base em similaridades químicas – *Similarity Search* - e capaz de produzir uma seriação dos eventuais resultados em função de um índice de semelhança. A implementação de tal procura por similaridades químicas ficaria então unicamente dependente da disponibilidade de ficheiros RXN que descrevessem quimicamente tanto as reacções a analisar como as reacções para comparação (neste caso concreto, as mesmas que serviram de base aos autores de *CC*).

---

### 2.2.1. DESENVOLVIMENTO DE *SCRIPT* PARA GERAR FICHEIROS RXN

---

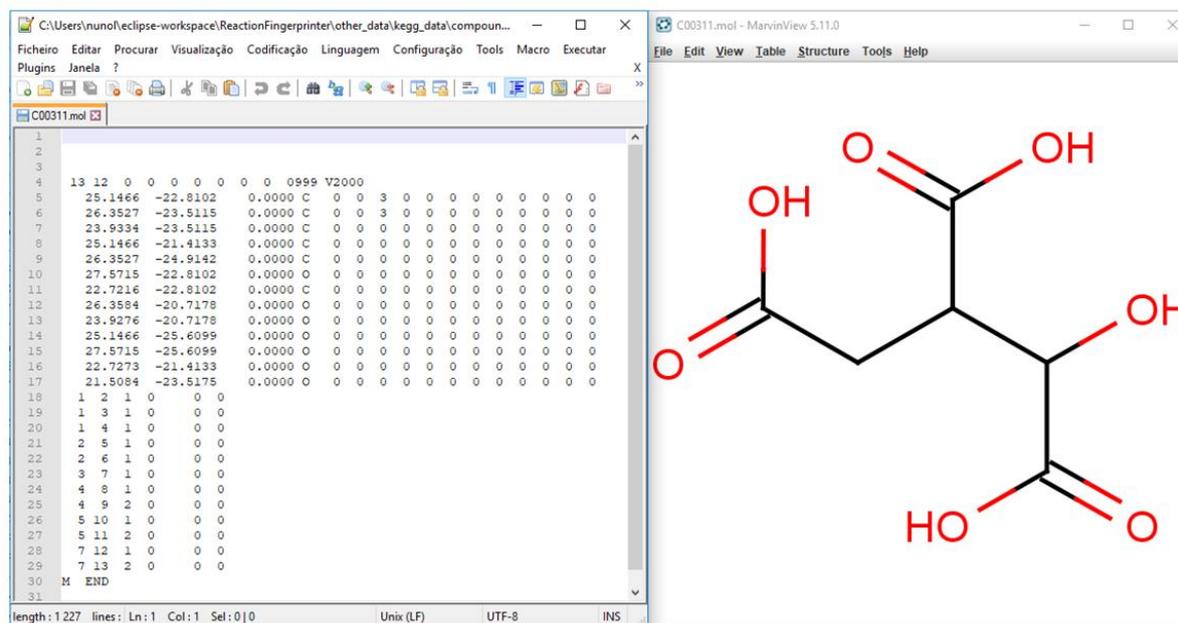
Um dos requisitos deste trabalho residia na necessidade de dispor de ficheiros RXN relativos a todas as reacções envolvidas nas análises a realizar – quer para as reacções da base de dados, para as quais as energias de Gibbs são já conhecidas, quer para aquelas para as quais se pretendia obter estimativas. Relativamente às reacções da base de dados para comparação, *TECR* não disponibiliza qualquer tipo de informação neste formato, pelo que seria necessário encontrar os respetivos ficheiros RXN noutra fonte ou desenvolver de raiz um novo método para a sua criação. De resto, esta segunda possibilidade acarretaria consigo uma capacidade acrescida de gerar ficheiros RXN para virtualmente qualquer reacção para a qual se queira obter estimativas, independentemente de constar ou não de alguma base de dados.

Foi neste contexto que se procedeu a um breve estudo sobre a estrutura dos ficheiros RXN e de quais seriam os requisitos para uma ferramenta que permita criá-los de raiz. De uma forma resumida, os ficheiros RXN – também conhecidos por *Reaction files (V2000)* – descrevem uma reacção em termos dos seus reagentes e produtos, cuja caracterização é por sua vez semelhante àquela que pode ser encontrada num ficheiro MDL.<sup>[53]</sup> Estes últimos consistem numa matriz que inclui: um bloco relativo aos átomos, '*Atom block*', com a enumeração de todos os átomos que constituem a molécula, descrevendo as suas coordenadas espaciais, cargas, electrões de valência e mapeamento das ligações químicas que os unem, entre outras propriedades; um bloco dedicado às ligações químicas, '*Bond block*', que descreve o tipo de cada ligação química (por exemplo: simples, dupla, tripla, aromática), a estereoquímica de cada ligação e a sua

topologia (cadeia ou anel); e um bloco respeitante a outras propriedades, 'Properties block', onde são assinaladas características tais como a presença de grupos -R ou radicais, massas isotópicas, entre outras. O ficheiro MDL é ainda encabeçado e encerrado por etiquetas identificativas do tipo de ficheiro.

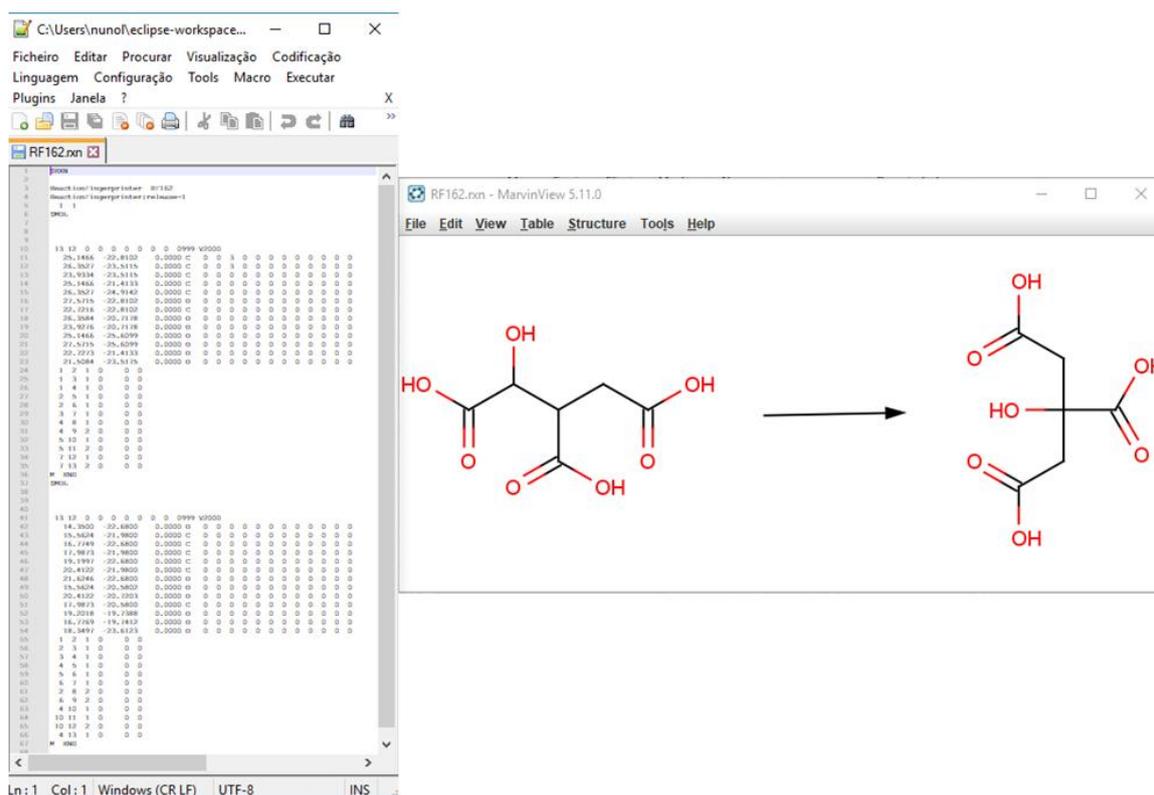
No caso dos ficheiros RXN, existe também uma etiqueta de cabeçalho. A seguir ao cabeçalho é possível inserir breves informações sobre o ficheiro, tais como uma referência identificativa da reação descrita ou sobre o programa através do qual o ficheiro foi criado. Segue-se uma linha com dois dígitos (ou números com mais do que um dígito, no caso de reações com um elevado número de reagentes e/ou produtos), relacionados, respetivamente, com o número de reagentes e o número de produtos da reação. A restante informação do ficheiro consiste na transcrição do ficheiro MDL de cada um dos reagentes ('Reactant block') e produtos ('Product block').

Tanto os ficheiros MDL como os ficheiros RXN são ainda passíveis de serem abertos através de interfaces que exibem a sua representação gráfica, como é o caso do *MarvinView*®. As imagens das Fig. 2.1. e Fig. 2.2. proporcionam uma demonstração visual da descrição feita sobre estes ficheiros.



**Fig. 2.1.** – Ficheiro MDL do isocitrato (KEGG Compound C00311 - [https://www.kegg.jp/dbget-bin/www\\_bget?cpd:C00311](https://www.kegg.jp/dbget-bin/www_bget?cpd:C00311)). À esquerda é possível visualizar a matriz que resume as informações sobre os átomos e ligações químicas que caracterizam este composto. À direita, eis a representação visual da molécula, obtida a partir da abertura do seu ficheiro MDL com o programa *MarvinView*®.

## Descrição das metodologias adoptadas e descrição das vias metabólicas estudadas



**Fig. 2.2.** – Ficheiro RXN relativo à reação de isomerização do isocitrato em citrato (*KEGG Reaction R01324* - [https://www.kegg.jp/dbget-bin/www\\_bget?rn:R01324](https://www.kegg.jp/dbget-bin/www_bget?rn:R01324)). À esquerda é visível a informação constante do ficheiro, incluindo a transcrição dos ficheiros MDL de cada um dos intervenientes na reacção. À direita, é possível visualizar a representação gráfica desta reacção através da abertura do ficheiro RXN com o programa *MarvinView*®.

Conhecidas as principais particularidades de ficheiros MDL e RXN, foi então possível construir um *script* em Python 2.7 capaz de ler uma lista de reacções no formato *KEGG* ( $C_{xxxxx} + \dots \rightleftharpoons C_{yyyyy} + \dots$ ; por exemplo, a isomerização do isocitrato em citrato: “C00311  $\rightleftharpoons$  C00158”) e de construir o respetivo ficheiro RXN para cada uma delas. Este *script* é ainda capaz de lidar com a eventual existência de um ID para cada reacção, localizado na mesma linha e antes da sua transcrição no formato *KEGG* (ID e reacção separados por uma tabulação; exemplo: “R01324<TAB>C00311  $\rightleftharpoons$  C00158”). Este *script* corresponde ao módulo *rxn\_builder.py* e compreende uma Classe *RxnBuilder*, concebida especialmente para a construção de ficheiros RXN para as reacções de *TECR*, e uma Classe *NewAnalysisRxnBuilder*, destinada à construção dos ficheiros RXN para todas as reacções a analisar através de *ReactionFingerprinter*.

Lidas as reacções, o algoritmo segue com o desdobramento de cada reacção nos blocos de reagentes e produtos, listando-os juntamente com os respetivos coeficientes estequiométricos. Conhecidos estes dados, é aberto um novo ficheiro, no qual são escritos o cabeçalho e contagem de reagentes e produtos. A listagem feita previamente

serve ainda para localizar os ficheiros MDL relativos a cada molécula e a sua transcrição para o ficheiro RXN. Neste passo, o ficheiro MDL de cada participante na reacção é obtido a partir da base de dados KEGG, que o disponibiliza na página respeitante à mesma molécula.

Até à data, não é conhecida nenhuma outra ferramenta igualmente capaz de, através de um modo simples e prático, construir ficheiros RXN para uma lista de reacções em moldes semelhantes aos que são aqui descritos. Por este motivo, a construção bem-sucedida de *rxn\_builder.py* pode ser considerada como um dos proveitos desta dissertação, podendo ainda beneficiar de alguns melhoramentos que visem a sua disponibilização pública em modo '*stand alone*'.

---

### 2.2.2. SIMILARITY SEARCH E REACTION FINGERPRINTS

---

A *Similarity search* é uma funcionalidade integrante de *JChem*® e visa a pesquisa de estruturas químicas com base na sua semelhança com uma estrutura de *query* (estrutura com base na qual a pesquisa é feita).<sup>[54]</sup> No caso concreto da pesquisa de reacções químicas, essa semelhança é calculada com base numa *Reaction fingerprint* - uma série de valores organizados de forma a constituir uma descrição pormenorizada dessa reacção.

Uma *Reaction fingerprint*, *RF*, engloba simultaneamente uma forma de comparar diretamente a estrutura dos reagentes e dos produtos de uma reacção (*Chemical fingerprints*, *CF*), bem como os aspetos relacionados com a transformação propriamente dita. A transformação pode ser descrita em três níveis distintos: o centro reativo (*coarse transformational similarity*) e o centro reativo alargado à vizinhança, numa extensão de uma (*medium-scale transformational similarity*) e duas (*strict transformational similarity*) ligações químicas.<sup>[55]</sup> Com base nestas considerações, a *RF* pode ser estruturada em oito segmentos diferentes:

- 1) *CF* dos reagentes;
- 2) *CF* dos produtos;
- 3) *CF* do centro reativo do lado dos reagentes;
- 4) *CF* de centro reativo do lado dos produtos;

## Descrição das metodologias adotadas e descrição das vias metabólicas estudadas

- 5) *CF* do centro reativo dos reagentes alargado à primeira ligação química;
- 6) *CF* do centro reativo dos produtos alargado à primeira ligação química;
- 7) *CF* do centro reativo dos reagentes alargado à segunda ligação química;
- 8) *CF* do centro reativo dos produtos alargado à segunda ligação química.

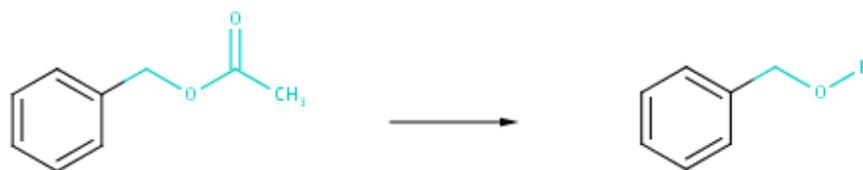
A extensão total de uma *RF* é de 2048 bits e os oito segmentos enumerados acima estão distribuídos segundo o padrão 512-512-128-128-128-128-256-256 (em número de bits).

Desta forma, *JChem*® encontra uma maneira conceptualmente simples para avaliar a similaridade entre reações químicas, aplicando a este caso técnicas idênticas às que são usadas para calcular a semelhança entre outros tipos de estruturas. Um pouco à imagem daquilo que foi idealizado pelos autores de *IGERS*, a abordagem escolhida para caracterizar uma transformação química reside principalmente na identificação dos átomos e ligações químicas que são modificados entre o lado dos reagentes e o lado dos produtos.

Na construção de uma *RF*, considera-se que um átomo sofre uma transformação se: uma ou mais das suas ligações químicas são alteradas; ou se, por exemplo, no caso de uma semi-reação, está presente apenas num dos lados da equação química e o seu átomo vizinho não sofre alterações. Uma ligação química é modificada quando aparece apenas num dos lados da reacção. Os átomos e as ligações químicas que são modificados constituem o centro reativo da reacção, sendo que este centro reativo é específico para um determinado tipo de reacção. A semi-reação ilustrada na **Fig. 2.3.** descreve estes critérios de uma forma mais intuitiva. As **Fig. 2.4.** e **Fig. 2.5.** transmitem também uma ideia mais clara acerca do alargamento do centro reativo da mesma semi-reação às primeira e segunda ligações químicas, respetivamente.



**Fig. 2.3.** – Identificação dos átomos e ligações químicas que são modificados, na perspetiva da construção de uma *Reaction Fingerprint*. Os átomos e ligações que participam na reacção estão assinalados a azul. Os números junto aos átomos do centro reativo funcionam como um mapeamento atómico que faz a correspondência entre o lado dos reagentes e o lado dos produtos. Este nível de análise de semelhança de transformação é também designado por *coarse transformational similarity*. [Imagem adaptada de [55]]



**Fig. 2.4.** – Alargamento do centro reativo à primeira ligação química, relativamente aos átomos modificados no decorrer da semi-reacção ilustrada na **Fig. 2.3.**. Os átomos e ligações visados estão coloridos a azul. Este nível de análise de semelhança de transformação é também designado por *medium-scale transformational similarity*. [Imagem adaptada de [55]]



**Fig. 2.5.** – Alargamento do centro reativo à segunda ligação química, relativamente aos átomos modificados no decorrer da semi-reacção ilustrada na **Fig. 2.3.**. Os átomos e as ligações visados estão coloridos a azul. Este nível de análise de semelhança é também designado por *strict transformational similarity*. [Imagem adaptada de [55]]

A condução de uma *Similarity search* pode ser aplicada a vários tipos de estruturas ou ficheiros. No caso concreto da comparação entre reacções químicas, a própria funcionalidade de *Similarity search* integra o processo de gerar uma *RF* quer para a estrutura a analisar – *query* – quer para a estrutura com a qual a primeira será comparada – *target*. Assim, após a criação dos ficheiros RXN de acordo com o explicitado na secção 2.2.1., a pesquisa de reacções por semelhança estrutural e transformacional ficava apenas dependente do estabelecimento de um ciclo capaz de conduzir uma sequência de comparações entre todas as *queries* e cada uma das reacções que formam a lista de *targets* para comparação, de acordo com um limiar de dissemelhança definido pelo utilizador.

Este ciclo de comparações faz uso de instruções passadas através da linha de comandos do sistema operativo e inclui a chamada de uma *Similarity Search* e a definição de uma série de argumentos: a nomeação da *query* (ficheiro RXN com a reacção a analisar) e do respetivo *target* (ficheiro RXN com a reacção com a qual a *query* é comparada)<sup>[54]</sup> e o limiar de dissemelhança definido<sup>[56, 57]</sup>. É ainda possível definir uma série de parâmetros adicionais, dependendo do objetivo da procura efectuada. No presente caso, optou-se por definir o tipo de índice a apresentar (nomeadamente o

## **Descrição das metodologias adoptadas e descrição das vias metabólicas estudadas**

índice de semelhança, havendo também a hipótese de seleccionar um índice de dissemelhança)<sup>[58]</sup>, bem como ativar a gravação dos resultados em ficheiros do tipo MRV, com a respetiva identificação do endereço e nome do ficheiro a gravar<sup>[58]</sup>. Para que melhor se perceba a estrutura completa da instrução, transcreve-se abaixo um exemplo para uma procura com base num limiar de 0,25:

```
"jcsearch -q caminho/para/o/ficheiro/de/query caminho/para/o/ficheiro/de/target -t:i:0.25 --similarity:s -f mrv -o caminho/para/o/ficheiro/com/o/resultado"
```

Neste exemplo, o segmento `"jcsearch -q"` é invocativo de uma nova *Similarity Search*. Depois da definição dos ficheiros de *query* e de *target*, é especificado o limiar de dissemelhança no valor de 0,25 (`"-t:i:0.25"`). Finalmente, define-se a apresentação dos resultados numa lógica de semelhança (ao invés de dissemelhança), através do argumento `"--similarity:s"` e transmite-se a instrução de gravação em ficheiro MRV (`"-f mrv"`), num local especificado pelo utilizador (`"-o caminho/para/o/ficheiro/com/o/resultado"`).

Os resultados da *Similarity Search* são então gravados em ficheiros, relativos a comparações entre pares *query-target*. Caso o índice de semelhança seja igual ou superior ao limiar definido, o ficheiro gravado inclui o seu valor. Este valor é obtido por uma ponderação (não especificada pela documentação da *ChemAxon*) entre coeficientes de Tanimoto relativos aos reagentes, aos produtos, e às *coarse, medium-scale* e *strict transformational similarities*.

Os cinco aspetos à luz dos quais a similaridade é calculada estão de acordo com os oito segmentos que compõem os 2048 bits da *RF* e que foram descritos mais acima. Assim, o coeficiente de Tanimoto dos reagentes é estabelecido com base no primeiro segmento da *RF*, enquanto que o coeficiente dos produtos é calculado a partir do segundo segmento. O coeficiente da *coarse transformational similarity* utiliza os terceiro e quarto segmentos e, na mesma lógica, os coeficientes da *medium-scale* e *strict similarity* são obtidos a partir, respetivamente, dos quinto e sexto e dos sétimo e oitavo segmentos.<sup>[55]</sup>

O coeficiente de Tanimoto é um índice de semelhança utilizado em vários contextos e, neste caso, é obtido através da expressão:

$$T = 1 - \frac{N_{ab}}{N_a + N_b - N_{ab}} \quad (\text{Equação 2.6.})$$

onde  $N_a$  corresponde ao número de bits fixado na  $RF_a$ ,  $N_b$  ao número de bits fixado na  $RF_b$  e  $N_{ab}$  ao número de bits em comum entre ambas as  $RF$ .

Note-se que, ao contrário do coeficiente de Tanimoto calculado em *IGERS* (secção 1.2.4.), neste caso observa-se uma subtração a partir da unidade, pelo que o coeficiente resultante diz respeito a um índice de dissemelhança. Independentemente disto, em *ReactionFingerprinter* este resultado é processado (subtraído à unidade), por forma a gerar um índice de semelhança entre reações, o que se afigura como bastante mais intuitivo. Finalmente, e para cada *query*, os coeficientes de Tanimoto obtidos para cada *target* são ordenados e apresentados num ficheiro final que contém um relatório completo acerca das semelhanças encontradas entre reações e respetivas estimativas para a energia livre de Gibbs. Além disto, é ainda gravado um outro ficheiro com a informação simplificada sobre as estimativas produzidas: a cada *query* é atribuído o valor da energia livre de Gibbs do *target* para o qual foi obtido um índice de semelhança mais elevado.

### 2.3. SELEÇÃO DA AMOSTRA DE 200 REAÇÕES METABÓLICAS

---

Proceder à aplicação e comparação de cada um dos métodos explicitados nas secções 2.1. e 2.2. implica a escolha de uma amostra com um determinado número de reações que possa ser testada por ambos. Ao ser analisado por ambos os métodos, tal conjunto de reações deverá permitir tirar conclusões acerca da cobertura proporcionada por cada método (ou por outras palavras, em que casos é que cada método não é capaz de gerar estimativas para a variação da energia livre de Gibbs), bem como perceber em que casos concretos é que as estimativas geradas apresentam uma maior discrepância entre si. No caso de *ReactionFingerprinter*, dado tratar-se de um método aqui desenvolvido de raiz, procurar-se-á ainda perceber qual a discrepância entre as estimativas geradas e alguns valores experimentais (ver próximas secções).

## Descrição das metodologias adoptadas e descrição das vias metabólicas estudadas

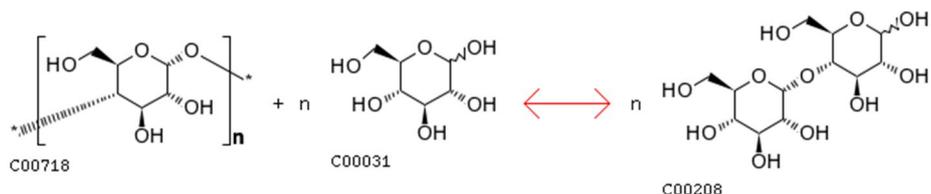
Esta selecção foi feita com base nas reacções disponíveis na base de dados KEGG e procurou, embora de uma forma não muito estruturada, englobar a maior variedade possível de tipos de reacções. Uma vez que *ReactionFingerprinter* depende em primeira instância da disponibilidade de ficheiros MDL (para a criação subsequente de ficheiros RXN), foi necessário excluir à partida reacções nas quais participam moléculas para as quais KEGG não apresenta o respectivo ficheiro MDL. Exemplo disso são reacções genéricas (que elucidam um tipo de mecanismo reacional e não a reacção concreta entre substratos específicos) (ver Fig. 2.6.: existem diferentes variantes de ferredoxinas e, neste caso em concreto, KEGG não disponibiliza uma estrutura devidamente detalhada para os compostos que intervêm na reacção) e reacções que envolvem ácidos nucleicos e/ou proteínas, moléculas estas demasiado grandes e complexas para serem representadas num ficheiro MDL (note-se que em alguns destes casos a própria estrutura da molécula não é completamente conhecida ou pode variar entre espécies e/ou indivíduos) (ver Fig. 2.7.). Há ainda o caso de reacções de polimerização – ou mecanismos idênticos –, para os quais não se conseguiu encontrar uma forma eficaz de lidar com os coeficientes estequiométricos (ver Fig. 2.8.).



**Fig. 2.6. – KEGG Reaction R00019.** A reacção representada refere-se à redução/oxidação de uma ferredoxina (C00138 – ferredoxina reduzida, C00139 – ferredoxina oxidada), catalisada pela enzima *E.C. 1.12.7.2.*. Descreve uma reacção genérica, cujos substratos não possuem uma estrutura que possa ser traduzida num ficheiro MDL. [Imagem adaptada de [52]]



**Fig. 2.7. – KEGG Reaction R00076.** A reacção representada refere-se à fosforilação/desfosforilação de ATP/ADP por acção enzimática (C02308 – fosforilase b [forma desfosforilada], C02307 – fosforilase a [forma fosforilada]), catalisada pela enzima *E.C. 2.7.11.19.*. Envolve a transformação daquela enzima entre uma forma ativa e inativa, ambas de elevada complexidade para poderem ser representadas em ficheiros MDL. [Imagem adaptada de [52]]



**Fig. 2.8. – KEGG Reaction R06069.** A reação representada refere-se ao alongamento de uma cadeia de polifosfato, catalisado pela enzima *E.C. 3.6.1.10.*. Dada a sua grande variabilidade de reação para reação, os coeficientes estequiométricos genéricos como os aqui representados tornam-se difíceis de tratar caso a caso e causam problemas na construção de ficheiros RXN para comparação de reações. [Imagem adaptada de [52]]

Note-se que *CC* também revela limitações na análise de reações que encaixam nas situações descritas: no caso das reacções genéricas ou que envolvam ácidos nucleicos ou sistemas proteicos, o desconhecimento de uma estrutura específica para uma ou mais moléculas, para a qual  $\Delta G_f^0$  possa ser estimada por *RC* ou decomposta em grupos por *GC*, invalida à partida a obtenção de uma estimativa; quando os coeficientes estequiométricos são genéricos (ou conduzem a inconsistências com o princípio de Lavoisier), *CC* é imediatamente interrompido, exigindo a correção das reações inseridas. Por estes motivos, acredita-se que esta abordagem está a excluir reações que vão para além das capacidades de estimativa de ambos os métodos e que por isso a sua ausência não interfere com a comparação que se pretende levar a cabo.

#### 2.4. SELEÇÃO DE UM GRUPO DE REAÇÕES COM $\Delta G_R^0$ CONHECIDA MAS AUSENTES DE TECR

A introdução de um método como *ReactionFingerprinter* exige também uma análise cuidada ao eventual desfasamento entre os seus resultados e a realidade. Para isso é necessário eleger um conjunto de reações para as quais  $\Delta G_r^0$  tenha sido determinada por trabalhos experimentais, ou para as quais a literatura apresente já estimativas sólidas. Igualmente importante é garantir que as reações incluídas nesse conjunto – ou, pelo menos, a maioria delas – não integram à partida a base de dados TECR. Caso um grande número de reações deste grupo fizesse parte de TECR, então encontrar-se-iam muitos índices de semelhança iguais a um, o que por sua vez eliminaria a lógica de obter estimativas por similaridade química (teríamos apenas a obtenção direta de um valor já conhecido à partida).

## ***Descrição das metodologias adoptadas e descrição das vias metabólicas estudadas***

A disponibilidade de dados termodinâmicos consistentes e que não estejam tabelados em TECR é escassa. Acima de tudo, TECR é já uma base de dados muito abrangente no que diz respeito à termodinâmica de reações enzimáticas e inclui a grande maioria daquelas para as quais são conhecidos dados obtidos em trabalhos experimentais. Esta escassez acentua-se especialmente quando se trata de encontrar informação normalizada relativamente ao efeito da temperatura, pH e força iónica. Além disto, e como já vimos na secção 2.3., é frequente que as descrições sobre reações metabólicas sejam pouco específicas em relação à participação de alguns co-factores e proteínas, referindo-se a estes por designações tais como “*um transportador de electrões oxidado*” ou “*uma flavoproteína reduzida*”. Tudo isto são situações que dificultam a aplicação de *ReactionFingerprinter* com o intuito com que se pretende fazê-lo nesta secção.

Contudo, e uma vez que a actual versão de *ReactionFingerprinter* é capaz de procurar todos os dados disponíveis em TECR para índices de semelhança iguais (isto é, provenientes de várias publicações acerca da mesma reacção), pode ser também interessante avaliar os seus resultados desta perspectiva. Por outras palavras, ainda que não se consigam normalizar determinados efeitos de forma consistente entre reações descritas por diferentes autores, é sempre possível acrescentar ao índice de semelhança uma lógica de similaridade das condições experimentais.

Em anos recentes registaram-se alguns esforços para determinar os dados termodinâmicos para um grande número de reações metabólicas em trabalhos que, em determinados aspetos, se distinguem dos dados disponíveis em TECR e do tipo de normalização levada a cabo pelos autores de *CC*. São disso exemplo uma tentativa de criar uma nova base de dados termodinâmicos para reações do ciclo dos ácidos tricarboxílicos (TCA), a glicólise a via das pentoses fosfato<sup>[59, 60]</sup>, bem como um estudo acerca do equilíbrio químico e cinética enzimática durante a atividade de células musculares<sup>[61]</sup>. Nota ainda para um interessante trabalho no sentido de investigar a termodinâmica de reações metabólicas em organismos termófilos e hipertermófilos.<sup>[62]</sup> Com base nestes exemplos, foi então possível escolher um pequeno grupo de dez reações que não constam de TECR e analisá-las por *ReactionFingerprinter*. A **Tabela 2.1**.

## Descrição das metodologias adotadas e das vias metabólicas estudadas

reúne as reações selecionadas, bem como as condições sob as quais a variação da sua energia livre de Gibbs foi calculada.

**Tabela 2.1.** – Reações metabólicas selecionadas para comparação dos valores de  $\Delta G_r$  estimados por *ReactionFingerprinter* com valores obtidos experimentalmente. (Legenda:  $\Delta G_r$  – variação da energia livre de Gibbs da reação (em *kJ/mol*); *T* (K) – temperatura (em graus Kelvin); *P* (bar) – pressão (em bar); *I* – força iônica do meio (em M).) [\*condições padrão (i.e.  $\Delta G_r^0$ )]

Reacção	$\Delta G_r$ (kJ/mol)	<i>T</i> (K)	<i>P</i> (bar)	pH	<i>I</i> (M)	Fonte
2-oxoglutarato + CoA + NADP <sup>+</sup> ⇌ CO <sub>2</sub> + NADPH + succinil-CoA	15,28*	298,15	-	-	0	[59, 60]
2 D-glucopiranoose 1-fosfato ⇌ D-glucose + D-glucose-1,6-bifosfato	-6,67	303,15	-	7,4	0,1	[61]
4 SO <sub>3</sub> <sup>2-</sup> + 2 H <sup>+</sup> ⇌ 3 SO <sub>4</sub> <sup>2-</sup> + H <sub>2</sub> S	-314,91	298	1	7	-	[62]
4 ácido fórmico ⇌ CH <sub>4</sub> + 3 CO <sub>2</sub> + 2 H <sub>2</sub> O	-177,54	298	1	7	-	[62]
ácido láctico + 2 NO <sub>3</sub> <sup>-</sup> ⇌ ácido acético + 2 NO <sub>2</sub> <sup>-</sup> + CO <sub>2</sub> + H <sub>2</sub> O	-327,54	298	1	7	-	[62]
2 ácido butanoico + 5 SO <sub>4</sub> <sup>2-</sup> + 10 H <sup>+</sup> ⇌ 5 H <sub>2</sub> S + 8 CO <sub>2</sub> + 8 H <sub>2</sub> O	-639,34	298	1	7	-	[62]
2 etanol + SO <sub>4</sub> <sup>2-</sup> + 2 H <sup>+</sup> ⇌ H <sub>2</sub> S + 2 ácido acético + 2 H <sub>2</sub> O	-188,20	298	1	7	-	[62]
2 propanol + SO <sub>4</sub> <sup>2-</sup> + 2 H <sup>+</sup> ⇌ 2 ácido propanoico + H <sub>2</sub> S + 2 H <sub>2</sub> O	-389,12	298	1	7	-	[62]
ácido acético + 8 Fe <sup>3+</sup> + 2 H <sub>2</sub> O ⇌ 8 Fe <sup>2+</sup> + 2 CO <sub>2</sub> + 8 H <sup>+</sup>	-495,25	298	1	7	-	[62]
ácido glicólico ⇌ glicolato + H <sup>+</sup>	21,88	298	1	7	-	[62]

## 2.5. DESCRIÇÃO DAS VIAS METABÓLICAS PARA ANÁLISE POR COMPONENT CONTRIBUTION E REACTION FINGERPRINTER

No âmbito de uma possível aplicação final à escala industrial, o foco dos métodos propostos neste trabalho deixará de estar em reacções bioquímicas isoladas. Será mais importante avaliar a termodinâmica de vias metabólicas compostas por várias reações, ou mesmo de mapas metabólicos alargados a várias vias interrelacionadas entre si. É

## ***Descrição das metodologias adotadas e descrição das vias metabólicas estudadas***

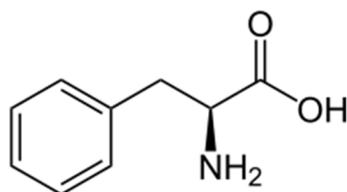
por isso sobejamente importante testar *CC* e *ReactionFingerprinter* sobre vias metabólicas com interesse científico e económico. Por este motivo, foram aqui selecionadas três vias metabólicas distintas, originalmente descritas para organismos diferentes, mas com potencial de aplicação extensível a vários modelos em biotecnologia e biologia de sistemas.

---

### 2.5.1. BIOSÍNTESE DE L-FENILALANINA

---

A fenilalanina é um aminoácido essencial, inserido no grupo dos aminoácidos aromáticos devido à presença de um anel de benzeno na sua cadeia lateral (**Fig. 2.9.**). O seu isómero L é biologicamente ativo, funcionando como precursor da L-tirosina e de vários neurotransmissores monoaminados, tais como a dopamina, a adrenalina e a noradrenalina, e integra ainda as vias de produção de melanina. O principal mercado para a L-fenilalanina reside na sua aplicação como precursor de *aspartame*®, um adoçante artificial e hipo-calórico. Outro grande motivo de interesse para o estudo das vias metabólicas nas quais este aminoácido exerce influência é a incidência de fenilcetonúria, uma doença genética relacionada com a incapacidade do organismo para processar a L-fenilalanina e que pode levar à sua acumulação até níveis tóxicos.<sup>[63]</sup>



**Fig. 2.9.** – Estrutura química bidimensional do aminoácido L-fenilalanina.

O estudo pormenorizado das vias metabólicas que conduzem à produção de L-fenilalanina, remonta já às décadas de 1970 e 1980. Desde então que são conhecidas a multiplicidade de vias metabólicas e sistemas de regulação que governam a sua produção em diversos organismos, motivando a procura de novos detalhes acerca da sua concreta influência e potencial de aplicação em biotecnologia.<sup>[64, 65]</sup> Este conhecimento tem sido usado em anos mais recentes para estimular a produção fermentativa de L-fenilalanina em quantidades cada vez mais economicamente viáveis.<sup>[66]</sup>

## *Descrição das metodologias adotadas e das vias metabólicas estudadas*

A sobre-expressão de enzimas do metabolismo de L-fenilalanina foi também já usada nas últimas décadas como abordagem para estudar a biossíntese de fenilpropanóides, com impacto ao nível da síntese de flavonoides e moléculas com potencial antimicrobiano.<sup>[67, 68]</sup> Já na década de 2000, o interesse direcionado para a síntese fermentativa de aditivos para a indústria alimentar esteve na génese de trabalhos voltados para a síntese de metabolitos aromáticos em que a fenilalanina funciona como precursor, tais como a vanilina, o benzaldeído e o cinamaldeído, nomeadamente através do fungo *Bjerkandera adusta*.<sup>[69]</sup> A aplicação na agricultura de derivados do metabolismo da L-fenilalanina, nomeadamente a classe dos glucosinolatos, enquanto protetores de culturas contra insetos e fungos é outro exemplo do potencial relacionado com as vias metabólicas deste aminoácido.<sup>[70]</sup>

No presente trabalho, a termodinâmica da biossíntese de L-fenilalanina será analisada do ponto de vista de uma via metabólica retirada da base de dados MetaCyc.<sup>[71]</sup> Neste caso, a via metabólica sob análise foi descrita para algumas bactérias e considera-se ter por precursores iniciais o fosfoenolpiruvato originário da glicólise e a D-eritrose 4-fosfato proveniente da via das pentoses fosfato. Esta via metabólica é composta por um total de dez reações:

- 1) fosfoenolpiruvato + D-eritrose 4-fosfato + H<sub>2</sub>O  $\rightleftharpoons$  3-desoxi-D-arabino-heptulose 7-fosfato + P<sub>i</sub>
- 2) 3-desoxi-D-arabino-heptulose 7-fosfato  $\rightleftharpoons$  3-deidroquinato + P<sub>i</sub>
- 3) 3-deidroquinato  $\rightleftharpoons$  3-deidroshikimato + H<sub>2</sub>O
- 4) 3-deidroshikimato + NADPH + H<sup>+</sup>  $\rightleftharpoons$  shikimato + NADP<sup>+</sup>
- 5) shikimato + ATP  $\rightleftharpoons$  shikimato 3-fosfato + ADP + H<sup>+</sup>
- 6) shikimato 3-fosfato + fosfoenolpiruvato  $\rightleftharpoons$  5-enolpiruvato-shikimato 3-fosfato + P<sub>i</sub>
- 7) 5-enolpiruvato-shikimato 3-fosfato  $\rightleftharpoons$  corismato + P<sub>i</sub>
- 8) corismato  $\rightleftharpoons$  pterinato
- 9) pterinato  $\rightleftharpoons$  3-fenil-2-oxopropanoato + CO<sub>2</sub> + H<sub>2</sub>O
- 10) 3-fenil-2-oxopropanoato + L-glutamato  $\rightleftharpoons$  L-fenilalanina + 2-oxoglutarato

---

2.5.2. VIA DE DEGRADAÇÃO DO GLICEROL E PRODUÇÃO DE BUTANOL

---

O butanol é um álcool de quatro carbonos (Fig. 2.10.) largamente aplicado enquanto solvente ou intermediário em vários processos químicos. É também conhecida a sua aplicação como biocombustível, com um poder de combustão muito próximo ao da gasolina. A produção de biobutanol é uma realidade há vários anos<sup>[72]</sup>, apesar de a sua toxicidade para organismos vivos ser demasiado elevada para que inicialmente pudessem ser obtidas quantidades economicamente viáveis. A elucidação dos passos metabólicos pelos quais o butanol pode ser obtido foi crucial no desenvolvimento de processos biotecnológicos capazes de superar esta limitação.<sup>[73]</sup>

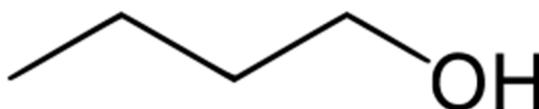


Fig. 2.10. - Estrutura química bidimensional do butanol.

O glicerol é um subproduto resultante da transesterificação de óleos vegetais para produção de biodiesel. (Estima-se que por cada 9 kg de biodiesel produzido por esta via é também obtido 1 kg de glicerol.) A sua utilização como matéria-prima noutras aplicações pode assumir também um relevo económico que está longe de ser negligenciável.<sup>[74, 75, 76]</sup> A via metabólica aqui explorada à luz dos métodos apresentados poderá num futuro próximo constituir uma forma científica e economicamente viável de obter um importante biocombustível a partir de um subproduto excedentário da indústria química.

Neste contexto, escolheu-se uma via metabólica descrita para bactérias do género *Clostridium*, disponível na base de dados MetaCyc.<sup>[77]</sup> É composta por um total de quinze reações distintas, com a particularidade de as últimas seis exigirem que todas as anteriores ocorram em duplicado para que no final ocorra a produção de uma molécula de butanol:

- 1) glicerol + NAD<sup>+</sup> ⇌ dihidroxiacetona + NADH + H<sup>+</sup>
- 2) dihidroxiacetona + ATP ⇌ glicerona fosfato + ADP + H<sup>+</sup>
- 3) glicerona fosfato ⇌ D-gliceraldeído 3-fosfato
- 4) D-gliceraldeído 3-fosfato + NAD<sup>+</sup> + P<sub>i</sub> ⇌ 3-fosfo-D-gliceroil-fosfato + NADH + H<sup>+</sup>

### Descrição das metodologias adotadas e das vias metabólicas estudadas

- 5) 3-fosfo-D-gliceroil-fosfato + ADP  $\rightleftharpoons$  3-fosfo-D-glicerato + ATP
- 6) 3-fosfo-D-glicerato  $\rightleftharpoons$  2-fosfo-D-glicerato
- 7) 2-fosfo-D-glicerato  $\rightleftharpoons$  fosfoenolpiruvato + H<sub>2</sub>O
- 8) fosfoenolpiruvato + ADP + H<sup>+</sup>  $\rightleftharpoons$  piruvato + ATP
- 9) piruvato + 2 ferredoxina oxidada (complexo [Fe-S]) + CoA  $\rightleftharpoons$  acetil-CoA + CO<sub>2</sub> + 2 ferredoxina reduzida (complexo [Fe-S]) + H<sup>+</sup>
- 10) 2 acetil-CoA  $\rightleftharpoons$  acetoacetil-CoA + CoA
- 11) acetoacetil-CoA + NADH + H<sup>+</sup>  $\rightleftharpoons$  (S)-3-hydroxibutanoil-CoA + NAD<sup>+</sup>
- 12) (S)-3-hydroxibutanoil-CoA  $\rightleftharpoons$  crotonoil-CoA + H<sub>2</sub>O
- 13) crotonoil-CoA + flavoproteína reduzida  $\rightleftharpoons$  butanoil-CoA + flavoproteína oxidada + H<sup>+</sup>
- 14) butanoil-CoA + NADPH + H<sup>+</sup>  $\rightleftharpoons$  butan-1-al + CoA + NADP<sup>+</sup>
- 15) butan-1-al + NADPH + H<sup>+</sup>  $\rightleftharpoons$  butan-1-ol + NADP<sup>+</sup>

---

#### 2.5.3. BIOSÍNTESE DE VIOLACEÍNA

---

O interesse na biossíntese da violaceína (Fig. 2.11.) remonta pelo menos ao início da década de 1990, com a sugestão de que este pigmento azulado, encontrado em várias bactérias, mas com especial frequência no género *Chromobacterium*, teria propriedades antibióticas contra diversos organismos. Além destas, especulava-se sobre um eventual efeito protetor contra a radiação visível e UV, bem como acerca de um possível mecanismo de remoção de triptofano e outros compostos derivados de indole.<sup>[78]</sup> Mais tarde apareceriam descrições de efeitos imunomoduladores, analgésicos e antipiréticos, bem como uma forte atividade contra células cancerígenas de vários tipos.<sup>[79, 80, 81, 82]</sup>

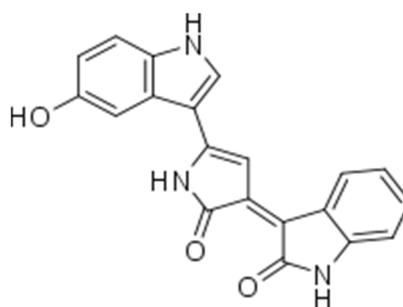


Fig. 2.11. - Estrutura química bidimensional da violaceína.

### ***Descrição das metodologias adoptadas e descrição das vias metabólicas estudadas***

A via metabólica aqui analisada para a síntese da violaceína está disponível na base de dados MetaCyc e provém de bactérias do género *Chromobacterium*.<sup>[83]</sup> Esta via tem início na transformação do L-triptofano e ramifica-se em vários pontos para a produção de diferentes moléculas. Contudo, numa lógica de obter uma cadeia simples - com início no triptofano e conclusão na formação de violaceína - esta via pode ser resumida em seis reações sequenciais:

- 1)  $\text{L-triptofano} + \text{O}_2 \rightleftharpoons \text{2-imino-3-(indol-3-il)propanoato} + \text{H}_2\text{O}_2 + \text{H}^+$
- 2)  $2 \text{ 2-imino-3-(indol-3-il)propanoato} + \text{H}_2\text{O}_2 \rightleftharpoons \text{N-[2-(carboxilatoamino)-1,2-bis(1H-indol-3-il)etil]carbamato} + 2 \text{ H}_2\text{O}$
- 3)  $\text{N-[2-(carboxilatoamino)-1,2-bis(1H-indol-3-il)etil]carbamato} + 2 \text{ H}^+ \rightleftharpoons \text{protodeoxiviolaceinato} + \text{NH}_4^+ + \text{CO}_2$
- 4)  $\text{protodeoxiviolaceinato} + \text{NAD(P)H} + \text{O}_2 + \text{H}^+ \rightleftharpoons \text{protoviolaceinato} + \text{NAD(P)}^+ + \text{H}_2\text{O}$
- 5)  $\text{protoviolaceinato} + \text{NAD(P)H} + \text{O}_2 + \text{H}^+ \rightleftharpoons \text{violaceinato} + \text{NAD(P)}^+ + \text{H}_2\text{O}$
- 6)  $\text{violaceinato} + \text{O}_2 + \text{H}^+ \rightleftharpoons \text{violaceína} + \text{CO}_2 + \text{H}_2\text{O}$

### **3. RESULTADOS E DISCUSSÃO**

---



Neste capítulo são apresentados e discutidos os resultados da aplicação dos métodos explorados nesta dissertação e descritos em detalhe no **Capítulo 2**. São aqui apresentados os resultados de *CC* e *ReactionFingerprinter* para uma amostra comum de duzentas reações da base de dados KEGG, que permitem acima de tudo comparar a abrangência proporcionada por cada um deles. São também apresentadas e comentadas as estimativas obtidas para um conjunto menor de reações, destinado a avaliar a discrepância entre os valores estimados por *ReactionFingerprinter* e aqueles que foram previamente calculados por outros autores. Há ainda lugar a uma exposição dos resultados obtidos para três outros conjuntos distintos de reações e que correspondem a três vias metabólicas completas, analisadas por ambos os métodos.

### 3.1. AVALIAÇÃO DA ABRANGÊNCIA DE *COMPONENT CONTRIBUTION* E *REACTION FINGERPRINTER*

---

A abrangência proporcionada por *CC* e *ReactionFingerprinter* é aqui referida no sentido do número e variedade de reações bioquímicas que são passíveis de ser analisadas por cada um dos métodos individualmente, com reflexo na produção de estimativas válidas para a correspondente variação na energia livre de Gibbs. Para avaliar este aspeto, foram selecionadas duzentas reacções que constam da base de dados KEGG e que foram escolhidas de acordo com os critérios explicitados na secção 2.3.. A abrangência de cada método é ainda explorada em termos da sua complementaridade mútua.

A **Tabela 3.1.** enumera as reações incluídas neste conjunto, bem como as energias de Gibbs estimadas através da aplicação de *CC*, associadas a uma estimativa do erro inerente ao seu cálculo e cujo cálculo é devidamente explicitado pelos autores do método aquando da sua publicação<sup>[20]</sup>.

## Resultados e discussão

**Tabela 3.1.** – Estimativas para a energia livre de Gibbs,  $\Delta G_r^\theta$ , obtidas por *Component Contribution* para a amostra de 200 reações da base de dados KEGG. *Component Contribution* calcula ainda o erro padrão,  $Err^{Std}$ , associado a cada estimativa.

KEGG ID	KEGG Reaction	$\Delta G_r^\theta$ (kJ/mol)	$Err^{Std}$ (+/- kJ/mol)
R00008	C06033 <=> 2 C00022	15.1	4.6
R00043	C04547 + C00007 <=> 2 C00755	-367.4	6.5
R00059	C01255 + C00001 <=> 2 C02378	2.0	3.3
R00075	3 C00019 <=> 3 C00170 + C05324	109.4	619806.4
R00085	C00002 + 2 C00001 <=> C00020 + 2 C00009	71.6	1.2
R00130	C00002 + C00882 <=> C00008 + C00010	47.5	5.8
R00251	C00002 + C01879 + 2 C00001 <=> C00008 + C00009 + C00025	32.9	6.2
R00269	C00940 + C00001 <=> C00026 + C00014	2.8	4.7
R00327	C02655 + C00001 <=> C00031 + C00033	12.7	4.6
R00348	C02362 + C00001 <=> C00036 + C00014	-4.5	5.0
R00470	C01127 <=> C00022 + C00048	16.6	3.0
R00658	C00631 <=> C00074 + C00001	-4.1	0.6
R00807	C00089 + C00016 <=> C05731 + C01352	-53.3	8.9
R00828	C01988 + C00007 + 2 C00005 + 2 C00080 <=> C00090 + C00088 + 2 C00006 + C00001	-763.5	277185.9
R00913	C03065 + C00001 <=> C00099 + C00086	-63.8	15.2
R00927	C00100 + C00024 <=> C00010 + C03344	32.1	5.6
R00998	C02488 + C00010 <=> C00024 + C00001 + C00109	-22.2	7.6
R01085	C02514 + C00001 <=> C00122 + C00022	-6.1	6.8
R01213	C02504 + C00010 <=> C00024 + C00141 + C00001	-17.1	2.6
R01334	C00988 + C00001 <=> C00160 + C00009	-31.7	4.1
R01518	C00631 <=> C00197	-4.1	0.7
R01576	C01187 <=> C00022 + C00216	18.5	4.5
R01635	C03233 <=> C00230 + C00011	-52.2	6.5
R01748	C00631 + C00001 <=> C00258 + C00009	-21.4	1.3
R01821	C00718 + C00009 <=> C00718 + C00103	0.0	0.0
R01933	C00322 + C00010 + C00003 <=> C00527 + C00011 + C00004 + C00080	-28.2	7.3
R01937	C00322 + C00014 + C00003 <=> C02220 + C00004 + C00080 + C00001	128.5	6.4
R02121	C00372 + C00718 <=> C00372 + C00718	0.0	0.0
R02160	C01617 + C00026 + C00007 <=> C00389 + C00042 + C00011 + C00001	-197.4	277185.9
R02177	C00400 + C00006 <=> C00843 + C00005 + C00080	52.1	3.1
R02232	C04316 <=> C00024 + C00414	-10.9	3.3
R02303	C00002 + C00448 <=> C00008 + C03115	2.6	2.5
R02340	C03506 <=> C00463 + C00118	19.2	5.8
R02505	3 C00083 + C00540 <=> 4 C00010 + C01745 + 4 C00011	-308.7	26.7
R02540	C02505 + C00001 <=> C07086 + C00014	-9.5	4.9
R02611	C05853 + C00003 <=> C00601 + C00004 + C00080	65.5	2.0
R02613	C05332 + C00007 + C00001 <=> C00601 + C00014 + C00027	-94.5	9.0
R02718	C00002 + C00133 + C00653 <=> C00020 + C00013 + C04260	0.0	0.0
R02722	C00065 + C03506 <=> C00078 + C00118 + C00001	-30.6	6.2
R02892	C03588 + C00006 <=> C00762 + C00005 + C00080	59.6	15.1
R02893	C05469 + C00006 <=> C00762 + C00005 + C00080	59.6	15.1

## Resultados e discussão

R02945	C01766 + C00003 <=> C00809 + C00004 + C00080	48.7	3.0
R03023	C00870 + C00007 + C00004 + C00080 <=> C02235 + C00003 + C00001	-483.7	7.0
R03024	C03360 + C00001 <=> C00870 + C00009	0.0	0.0
R03035	C00002 + C01134 <=> C00013 + C00882	-22.8	10.1
R03063	C00916 + C00026 <=> C04712 + C00217	-8.1	5.6
R03083	C04691 <=> C00944 + C00009	-132.8	9.8
R03126	C00974 + C00026 + C00007 <=> C05903 + C00042 + C00011 + C00001	-197.4	277185.9
R03180	C03078 + C00001 <=> C01035 + C00014	-11.1	4.4
R03207	C01060 + C00026 <=> C01244 + C00025	-2.7	4.3
R03249	C01110 <=> C03564 + C00001	-8.5	10.5
R03250	C01110 + C00086 <=> C03771 + C00001	57.7	11.0
R03264	C01126 + C00006 <=> C03461 + C00005 + C00080	68.9	2.0
R03265	C01126 <=> C03220	0.0	0.0
R03307	C01278 <=> C01163	35.9	7.0
R03325	C04518 + C00006 <=> C01176 + C00005 + C00080	66.4	2.0
R03336	C03672 + C00003 <=> C01179 + C00004 + C00080	118.8	8.9
R03350	C04478 + C00001 <=> C01187 + C00009	-13.7	2.4
R03351	C00063 + C01187 <=> C00013 + C04121	264.8	138592.9
R03378	C00044 + C16741 <=> C00035 + C03366	34.4	3.7
R03448	C04277 + C00001 <=> C01255	-21.4	16.0
R03470	C01278 <=> C03586 + C00011	-118.4	10.4
R03567	C02657 <=> C01467 + C00011	-52.2	6.5
R03614	C01575 + C00003 <=> C04351 + C00004 + C00080	87.1	98000.0
R03638	C01617 + C00007 + C00005 + C00080 <=> C03052 + C00006 + C00001	-487.1	7.0
R03723	C03910 <=> C01802 + C00084	22.8	6.8
R03761	C01892 + C00004 + C00080 <=> C04371 + C00003	-45.9	2.5
R03782	C02289 <=> C01951	-78.8	6.6
R03908	C02930 <=> C02295	3.8	5.8
R03954	C02470 + C00001 <=> C05645	116.5	10.1
R03959	C04105 <=> C02480	50.9	2.6
R04031	C02730 + C00001 <=> C05817	12.7	11.7
R04059	C04167 + 2 C00006 <=> C02807 + 2 C00005 + 2 C00080	104.2	6.1
R04133	C04052 <=> C03063 + C00011	-61.0	6.1
R04141	C03837 + C00001 <=> C03081 + C00033	-26.5	6.1
R04151	C04468 + C00003 <=> C03164 + C00004 + C00080	-29.1	12.4
R04188	C03284 + C00026 <=> C06002 + C00025	-1.4	3.7
R04209	C04751 <=> C03373 + C00011	-42.4	8.5
R04229	C03474 + 2 C00005 + 2 C00080 <=> C04050 + 2 C00006 + C00001	0.0	0.0
R04259	C04522 <=> C03585	35.9	7.0
R04260	C04105 <=> C03586	-51.1	7.1
R04277	C03671 + C00001 <=> C04434	-196.1	196000.0
R04320	C00470 + C00001 <=> C00470 + C00333	0.0	0.0
R04343	C00617 + C00470 <=> C00015 + C00470	0.0	0.0
R04367	C03995 <=> C02126 + C04815	0.0	0.0
R04369	C00002 + C00133 + C03999 <=> C00008 + C00009 + C04457	13.5	4.8

## Resultados e discussão

R04371	C05662 <=> C04002 + C00001	-1.6	5.7
R04376	C04044 + C00007 <=> C04479	-214.8	10.3
R04378	C04677 + C00013 <=> C04051 + C00119	-8.0	7.1
R04439	C04272 + C00006 <=> C06010 + C00005 + C00080	66.4	2.0
R04441	C04272 <=> C00141 + C00001	-35.5	2.7
R04455	C04330 + C00282 <=> C04377 + C00080	0.5	8.7
R04520	C00353 + C04590 <=> C00013 + C04638	20.6	5.4
R04559	C04823 <=> C00122 + C04677	5.1	5.8
R04580	C00002 + C04722 + C00010 <=> C00020 + C00013 + C04760	31.5	2.9
R04602	C04798 + C00007 <=> C05155	0.0	0.0
R04623	C04875 + C00006 <=> C04883 + C00005 + C00080	52.1	3.1
R04809	C05447 + C00001 <=> C05448	3.2	3.7
R04812	C05450 + C00003 <=> C05467 + C00004 + C00080	63.0	1.9
R04813	C05450 <=> C05460 + C00001	-3.2	3.7
R04891	C05583 + C00003 + C00001 <=> C05584 + C00004 + C00080	46.5	2.2
R04987	C05810 + C00007 + C00005 + C00080 <=> C05811 + C00006 + C00001	-487.1	7.0
R05078	C06033 <=> C06035 + C00001	3.2	3.7
R05092	C06089 <=> C06090 + C00013	0.0	0.0
R05153	C06328 + C00007 <=> C06329	-212.2	196000.0
R05258	C00472 + C00088 + C00006 + C00001 <=> C00870 + C00007 + C00005 + C00080	445.1	196000.0
R05347	C07216 + C00003 <=> C07209 + C00004 + C00080	65.5	2.0
R05356	C06593 + C00001 <=> C02378	6.7	8.2
R05391	C03585 <=> C04431 + C01327	0.0	0.0
R05507	C02378 + C00026 <=> C06102 + C00025	-1.4	3.7
R05601	C06723 + C00001 <=> C06714	-0.7	2.9
R05613	C04217 + C00129 <=> C01230 + C00013	-30.1	5.7
R05637	C11436 <=> C11453 + C00055	1815.5	196000.0
R05639	C05108 + C00006 <=> C11455 + C00005 + C00080	59.6	15.1
R05697	C11500 + C00007 + C00001 <=> C11501 + C00027	-129.3	8.5
R05773	C07303 <=> C00755 + C00024	-0.1	2.2
R05790	C11516 + C00001 <=> C11544 + C00009	-13.7	2.4
R05791	C10517 + C00001 <=> C11546 + C00009	-13.7	2.4
R06117	C11409 + C00003 <=> C11383 + C00004 + C00080	48.7	3.0
R06301	C11901 <=> C11878 + C00013	-114.9	35.2
R06302	C11901 <=> C11879 + C00013	-114.9	35.2
R06303	C11901 <=> C11880 + C00013	-97.5	34.0
R06378	C11397 + C00003 <=> C11398 + C00004 + C00080	48.7	3.0
R06380	C11398 <=> C11401	0.0	0.0
R06382	C11399 + C00003 <=> C11401 + C00004 + C00080	48.7	3.0
R06383	C11413 + C00003 <=> C11415 + C00004 + C00080	48.7	3.0
R06417	C02485 + C00005 + C00080 <=> C11951 + C00006	-10.4	6.6
R06420	C09893 + C00005 + C00080 <=> C11952 + C00006	-27.8	6.5
R06532	C05906 + C00005 + C00080 <=> C06562 + C00006 + C00001	-97.3	2.6
R06533	C05909 + C00005 + C00080 <=> C12127 + C00006 + C00001	-97.3	2.6
R06539	C02906 + C00026 + C00007 <=> C10107 + C00042 + C00011 + C00001	-197.4	277185.9

## Resultados e discussão

R06541	C12128 + 2 C00006 <=> C05904 + 2 C00005 + C00080	623.1	438269.3
R06542	C09727 + 2 C00006 <=> C05905 + 2 C00005 + C00080	623.1	438269.3
R06543	C12136 + 2 C00006 <=> C05908 + 2 C00005 + C00080	623.1	438269.3
R06583	C05619 + C00010 + C00002 <=> C12203 + C00020 + C00013	31.5	2.9
R06615	C03648 + C00005 + C00080 <=> C09320 + C00006 + C00001	-97.3	2.6
R06784	C11588 + C00003 <=> C04044 + C00004 + C00080	15.9	11.5
R06786	C11457 + C00007 + C00004 + C00080 <=> C04044 + C00001 + C00003	-483.7	7.0
R06857	C12837 + C00003 <=> C05618 + C00004 + C00080	2.2	196000.0
R06866	C05313 <=> C05802 + C00011	-52.2	6.5
R06892	C10700 + C00005 + C00080 + C00007 <=> C13636 + C00006 + C00001	-475.9	7.0
R06957	C13455 + C00001 + C00007 <=> C06082 + C00027	-129.3	8.5
R06975	C00058 + C00002 + C04677 <=> C00008 + C00009 + C04734	-2.8	7.0
R06987	C00109 + C00010 <=> C00100 + C00058	-2.2	9.4
R06991	C14721 <=> C12247	0.0	0.0
R07136	C11537 + C00003 <=> C05528 + C00004 + C00080	63.0	1.9
R07137	C11537 + C00006 <=> C05528 + C00005 + C00080	66.4	2.0
R07271	C00235 + C01263 <=> C00013 + C15509	-31.3	6.4
R07272	C00235 + C01263 <=> C00013 + C15510	-31.3	6.4
R07404	C00002 + C03373 + C00288 <=> C00008 + C00009 + C15667	21.7	10.7
R07430	C15789 + C00007 + C00005 <=> C15790 + C00006 + C00001	-443.2	6.1
R07433	C10434 + C00010 <=> C00323 + C00493	-1.7	4.6
R07633	C16069 <=> C00022 + C11481	0.0	0.0
R07649	C00448 <=> C16141 + C00013	-214.4	20.3
R07726	C10200 <=> C10205	0.0	0.0
R07727	C16234 + 2 C00005 + 2 C00080 <=> C16235 + 2 C00006 + C00001	-253.7	277185.9
R07736	C16188 <=> C16189	0.0	0.0
R07788	C15797 <=> C16251	0.0	0.0
R07801	C16263 + C00007 <=> C16264	-214.0	10.9
R07839	C04434 <=> C04324	0.0	0.0
R07984	C04226 + 2 C00001 <=> C16159 + C00014	95.0	8.0
R07987	C00540 + 3 C00083 <=> C16404 + 4 C00010 + 3 C00011	-62.5	196000.0
R08058	C01217 + C00067 <=> C04377 + C00001	-34.8	6.5
R08059	C04377 + C00006 <=> C04330 + C00005	10.8	6.6
R08093	C16468 + C00001 <=> C16469	-0.7	2.9
R08466	C08301 <=> C16719	0.0	0.0
R08625	C17215 + C00003 <=> C17216 + C00011 + C00004 + C00080	2.0	6.2
R08629	C17219 + C00003 <=> C17220 + C00011 + C00004 + C00080	2.0	6.2
R08734	C15613 <=> C17343	0.0	0.0
R08766	C03672 <=> C00811 + C00001	0.7	2.9
R08831	C05587 + C08538 <=> C17754 + C00001	26.0	98000.0
R08837	C17758 <=> C08538 + C00001	-155.7	17.2
R08842	C17733 <=> C17735	0.0	0.0
R08853	C16269 + C00005 + C00080 + C00007 <=> C17955 + C00006 + C00001	0.0	0.0
R08992	C00353 + 2 C00001 <=> C09183 + C00013	-128.2	196000.0
R09043	C18132 <=> C18133	0.0	0.0

## Resultados e discussão

R09059	C02139 + C00001 <=> C18154	0.0	0.0
R09111	C11901 <=> C18211 + C00013	-13.4	33.3
R09119	C06089 <=> C18226 + C00013	-72.4	29.8
R09120	C06089 <=> C11877 + C00013	-13.4	33.3
R09121	C06089 <=> C18228 + C00013	-13.4	33.3
R09214	C18302 + C00007 <=> C18303	-190.0	392000.0
R09283	C06028 <=> C18323	0.0	0.0
R09292	C20238 <=> C18324	0.0	0.0
R09358	C18796 <=> C02845 + C01326	96.2	138592.9
R09487	C16143 + C00001 <=> C16286 + C00207	23.0	196000.0
R09545	C19675 + 2 C00003 + C00001 <=> C11537 + 2 C00004 + 2 C00080	112.0	2.6
R09614	C00448 <=> C19740 + C00013	-321.1	619806.4
R09664	C19791 + C00001 <=> C19792 + C00009	-13.7	2.4
R09751	3 C00088 <=> 2 C00533 + C00244 + C00001	46.8	21.8
R09783	C19892 + C00003 + C00001 <=> C08607 + C00004 + C00080	46.5	2.2
R09835	C19973 + C00001 <=> C19974	67.5	277185.9
R09874	C00353 + C00001 <=> C20158 + C00013	-210.7	196000.0
R09967	C20270 <=> C00353 + C00001	143.1	196000.0
R10072	C16269 + 2 C00005 + 2 C00080 + 2 C00007 <=> C17954 + 2 C00006 + 3 C00001	0.0	0.0
R10111	C20367 <=> C20366 + C00088	-112.8	196000.0
R10133	C14531 + C00001 <=> C20382 + C00088	39.9	339482.0
R10306	3 C00019 + C04441 <=> 3 C00021 + C01573	0.0	0.0
R10350	C20525 + C00007 + C00005 + C00080 <=> C20526 + C00006 + C00001	-463.7	8.6
R10829	C20887 + C00009 <=> C00140 + C00636	2.2	5.5

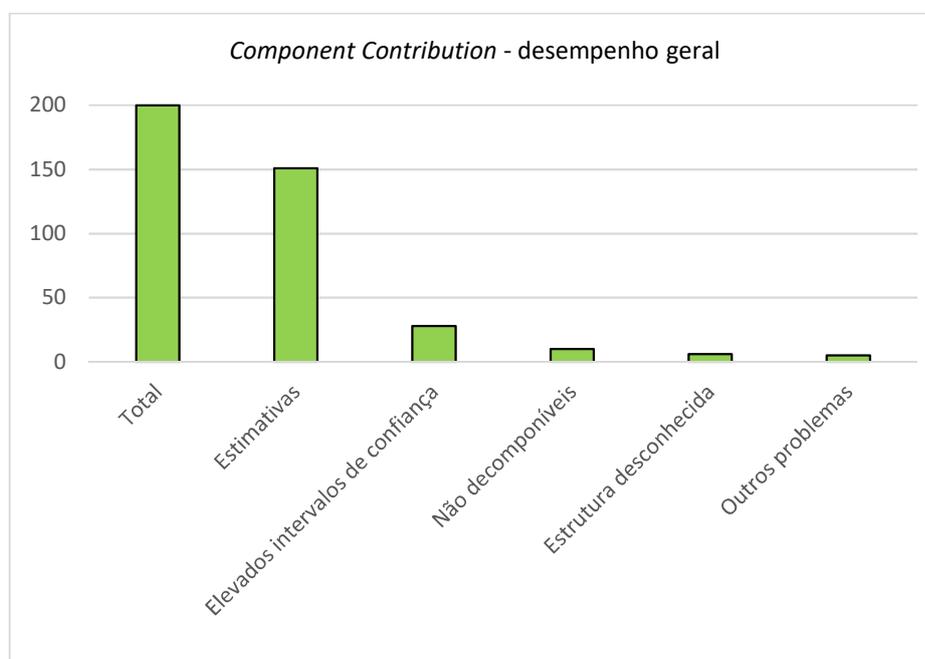
Num cômputo geral, das duzentas reações analisadas por *CC*, vemos que 27 originaram uma estimativa de zero para a energia livre de Gibbs. Dessas 27, 6 correspondem a reações de isomerização (R03265, R06380, R07839, R08734, R08842 e R09283), enquanto que as restantes dizem respeito a casos que *CC* não é capaz de processar: 6 reações envolvem moléculas com estruturas desconhecidas (R01821, R02121, R02718, R04320, R04343 e R10306), 10 incluem moléculas com estruturas não decomponíveis à luz de *CC* (R03024, R04249, R04602, R05092, R05391, R07633, R07736, R08853, R09059 e R10072) e 5 onde é identificado outro tipo de problema, como por exemplo a existência de um desequilíbrio químico (R06991, R07726, R07788, R08466, R09043). No total, estas 21 reações correspondem a 10,5% do total da amostra de 200 reações de KEGG.

Atendendo aos exemplos citados, não parece ser possível definir para *CC* zonas de ineficácia em termos de tipos ou mecanismos de reação. Em vez disso, a capacidade de processar reações parece estar mais relacionada com a estrutura dos compostos que nelas intervêm. Mais concretamente, a existência de grupos *-R* (grupos moleculares de estrutura não-especificada, sem interferência na transformação ocorrida), polímeros ou

estruturas de elevada complexidade parecem impedir sempre o processamento de uma reação por *CC*, independentemente do tipo de transformação ocorrida. O único caso em que é possível definir à partida a incapacidade de processamento está nas reações de polimerização, uma vez que envolvem sempre compostos com estruturas repetitivas, que são abreviadas de uma forma que *CC* não consegue interpretar. Ainda assim, o cerne do problema continua a estar nos compostos em si mesmos e não no tipo de reação.

Além das reações que *CC* não é capaz de processar, há ainda 28 reações para as quais não é possível produzir estimativas com um intervalo de confiança aceitável. Estas correspondem a 14% do total da amostra considerada. Estes casos específicos resultam da atribuição de uma margem de erro artificial sempre que os dados experimentais disponíveis não permitam validar de forma razoável a estimativa gerada e são detalhados na secção 3.1.5..

Somando todos estes casos, a cobertura proporcionada por *CC* estende-se a aproximadamente 74,5% das reações consideradas. Os 25,5% restantes incluem aquelas reações para as quais não foi possível obter estimativas ou para as quais não foi possível obter um intervalo de confiança suficientemente estreito. O gráfico da **Fig. 3.1.** transmite uma ideia geral do desempenho de *CC* para o conjunto de reações considerado.



**Fig. 3.1.** – Representação gráfica do desempenho geral de *CC* para a amostra de 200 reações de KEGG, ilustrando os parciais das reações para as quais foram geradas estimativas, bem como para aquelas para as quais não foi possível obter intervalos de confiança razoáveis, e ainda todas as que envolvem algum tipo de característica que impeça o seu correto processamento por este método.

## Resultados e discussão

A Tabela 3.2. resume os resultados obtidos através de *ReactionFingerprinter*, organizados em termos das reacções de *query* (amostra de 200 reacções de KEGG) e do *target* para o qual se obteve o maior índice de semelhança. As colunas seguintes indicam, respetivamente o valor estimado para a variação da energia livre de Gibbs, bem como o desvio-padrão associado a este valor, e o limiar mínimo de dissemelhança para o qual foi possível obter um emparelhamento *query-target*. Convém lembrar que o algoritmo de *ReactionFingerprinter* foi corrido três vezes para esta amostra, com a variação do limiar de dissemelhança entre 0,3, 0,25 e 0,2. O valor estimado para a variação da energia livre de Gibbs de cada reacção corresponde a uma média aritmética simples dos valores de  $\Delta G_r^0$  disponíveis em TECR para o *target* com melhor índice de semelhança. É frequente que o desvio-padrão associado aos valores de  $\Delta G_r^0$  seja de zero, o que normalmente corresponde aos casos para os quais TECR disponibiliza uma única medição de  $\Delta G_r^0$ , ou, menos frequentemente, existe um reduzido número de medições (geralmente duas ou três), todas com o mesmo valor de  $\Delta G_r^0$ .

Das 200 reacções analisadas com *ReactionFingerprinter*, foi possível obter estimativas para um total de 126, perfazendo uma cobertura de cerca de 63% da amostra estudada. Para a maioria destas, 116 no total (58% da amostra), foi ainda possível encontrar um emparelhamento *query-target* para um limiar de dissemelhança de 0,2. Ao alargar este limiar a 0,25 e 0,3, o total de emparelhamentos obtidos subiu, respectivamente, para 123 (7 reacções adicionais) e 126 (3 reacções adicionais).

Também aqui não é perceptível que a abrangência de *ReactionFingerprinter* esteja necessariamente dependente do tipo de transformação ocorrida, mas antes da existência de dados experimentais de reacções suficientemente parecidas com as que compõem a amostra. Note-se ainda que o alargamento do valor do limiar não conduziu a um aumento muito significativo do número de reacções abrangidas, pelo que não é garantido que maiores alargamentos proporcionem um grande aumento da cobertura (pelo menos tendo por base apenas as reacções de TECR).

**Tabela 3.2.** - Estimativas para a energia livre de Gibbs,  $\Delta_r G^\theta$ , obtidas por *ReactionFingerprinter* para a amostra de 200 reacções da base de dados KEGG. As estimativas são obtidas pela atribuição do valor médio das energias livres de Gibbs disponíveis em TECR para a reacção com maior índice de semelhança relativamente à reacção de *query*. A cada estimativa está ainda associado o desvio-padrão, *StdDev*, da medição e o maior índice de semelhança para o qual foi possível obter um emparelhamento *query-target*.

KEGG ID	KEGG Reaction	Melhor target	$\Delta_r G^\theta$ (kJ/mol)	StdDev	Índice de semelhança
R00008	C06033 <=> 2 C00022	C01127 <=> C00022 + C00048	10,85	1,33	0,879
R00043	C04547 + C00007 <=> 2 C00755				
R00059	C01255 + C00001 <=> 2 C02378	C00001 + C16512 <=> C00189 + C00249	22,52	0,00	0,842
R00075	3 C00019 <=> 3 C00170 + C05324	C00001 + C00021 <=> C00155 + C00212	34,74	0,00	0,884
R00085	C00002 + 2 C00001 <=> C00020 + 2 C00009	C00002 + C00033 <=> C00008 + C00227	10,29	1,57	0,980
R00130	C00002 + C00882 <=> C00008 + C00010	C00002 + C00882 <=> C00008 + C00010	47,47	0,00	1,000
R00251	C00002 + C01879 + 2 C00001 <=> C00008 + C00009 + C00025	C00002 + C00014 + C00025 <=> C00008 + C00009 + C00064	22,79	2,09	0,947
R00269	C00940 + C00001 <=> C00026 + C00014	C00048 + C00064 <=> C00037 + C00940	-16,22	0,00	0,778
R00327	C02655 + C00001 <=> C00031 + C00033	C06468 <=> C06464	3,33	0,00	0,750
R00348	C02362 + C00001 <=> C00036 + C00014				
R00470	C01127 <=> C00022 + C00048	C01127 <=> C00022 + C00048	10,85	1,33	1,000
R00658	C00631 <=> C00074 + C00001	C00631 <=> C00001 + C00074	-3,31	0,67	1,000
R00807	C00089 + C00016 <=> C05731 + C01352	C00016 <=> C01352	-77,55	0,00	0,962
R00828	C01988 + C00007 + 2 C00005 + 2 C00080 <=> C00090 + C00088 + 2 C00006 + C00001	C00006 + C03351 <=> C00005 + C03067	55,18	0,00	0,965
R00913	C03065 + C00001 <=> C00099 + C00086				
R00927	C00100 + C00024 <=> C00010 + C03344	2 C00024 <=> C00010 + C00332	25,68	2,29	0,993
R00998	C02488 + C00010 <=> C00024 + C00001 + C00109	C00001 + C00024 + C00036 <=> C00010 + C00158	10,09	1,64	0,993
R01085	C02514 + C00001 <=> C00122 + C00022				
R01213	C02504 + C00010 <=> C00024 + C00141 + C00001	C00001 + C00024 + C00036 <=> C00010 + C00158	10,09	1,64	0,993
R01334	C00988 + C00001 <=> C00160 + C00009				
R01518	C00631 <=> C00197	C00631 <=> C00197	-5,17	1,25	1,000
R01576	C01187 <=> C00022 + C00216	C01019 <=> C01721	5,48	0,00	0,800
R01635	C03233 <=> C00230 + C00011				
R01748	C00631 + C00001 <=> C00258 + C00009	C00631 <=> C00197	-5,17	1,25	0,836
R01821	C00718 + C00009 <=> C00718 + C00103	C00103 <=> C00092	-6,46	2,21	0,972

## Resultados e discussão

R01933	C00322 + C00010 + C00003 <=> C00527 + C00011 + C00004 + C00080	C00003 + C05268 <=> C00004 + C05269	67,04	3,39	0,967
R01937	C00322 + C00014 + C00003 <=> C02220 + C00004 + C00080 + C00001	C00001 + C00003 + C05662 <=> C00004 + C00288 + C00322	44,81	0,00	0,967
R02121	C00372 + C00718 <=> C00372 + C00718	C00492 + C01235 <=> C00137 + C01613	-3,48	0,04	0,905
R02160	C01617 + C00026 + C00007 <=> C00389 + C00042 + C00011 + C00001				
R02177	C00400 + C00006 <=> C00843 + C00005 + C00080	C00006 <=> C00005	22,59	0,00	0,982
R02232	C04316 <=> C00024 + C00414	C05268 <=> C00001 + C05271	1,03	0,00	0,964
R02303	C00002 + C00448 <=> C00008 + C03115	C00002 + C00008 <=> C00020 + C03483	52,06	0,00	0,933
R02340	C03506 <=> C00463 + C00118				
R02505	3 C00083 + C00540 <=> 4 C00010 + C01745 + 4 C00011	C01213 <=> C00683	0,00	0,00	0,905
R02540	C02505 + C00001 <=> C07086 + C00014				
R02611	C05853 + C00003 <=> C00601 + C00004 + C00080	C00003 + C00556 <=> C00004 + C00261	63,23	0,15	0,975
R02613	C05332 + C00007 + C00001 <=> C00601 + C00014 + C00027				
R02718	C00002 + C00133 + C00653 <=> C00020 + C00013 + C04260	C00002 + C00008 <=> C00020 + C03483	52,06	0,00	0,958
R02722	C00065 + C03506 <=> C00078 + C00118 + C00001	C00118 + C00463 <=> C03506	-19,17	0,00	0,908
R02892	C03588 + C00006 <=> C00762 + C00005 + C00080	C00003 + C00535 <=> C00004 + C00280	45,48	1,79	0,968
R02893	C05469 + C00006 <=> C00762 + C00005 + C00080	C00003 + C00535 <=> C00004 + C00280	45,48	1,79	0,968
R02945	C01766 + C00003 <=> C00809 + C00004 + C00080	C00003 + C00854 <=> C00004 + C00414	50,95	1,12	0,988
R03023	C00870 + C00007 + C00004 + C00080 <=> C02235 + C00003 + C00001	C00003 + C00556 <=> C00004 + C00261	63,23	0,15	0,952
R03024	C03360 + C00001 <=> C00870 + C00009				
R03035	C00002 + C01134 <=> C00013 + C00882	C00002 + C00882 <=> C00008 + C00010	47,47	0,00	0,989
R03063	C00916 + C00026 <=> C04712 + C00217	C00001 + C07761 <=> C02595 + C07756	12,03	0,00	0,805
R03083	C04691 <=> C00944 + C00009	C04442 <=> C00022 + C00118	13,09	0,34	0,822
R03126	C00974 + C00026 + C00007 <=> C05903 + C00042 + C00011 + C00001				
R03180	C03078 + C00001 <=> C01035 + C00014				
R03207	C01060 + C00026 <=> C01244 + C00025	C00026 + C00082 <=> C00025 + C01179	-1,20	0,04	0,808
R03249	C01110 <=> C03564 + C00001				
R03250	C01110 + C00086 <=> C03771 + C00001				
R03264	C01126 + C00006 <=> C03461 + C00005 + C00080	C00006 <=> C00005	22,59	0,00	0,988
R03265	C01126 <=> C03220				

## Resultados e discussão

R03307	C01278 <=> C01163				
R03325	C04518 + C00006 <=> C01176 + C00005 + C00080	C00003 + C00535 <=> C00004 + C00280	45,48	1,79	0,977
R03336	C03672 + C00003 <=> C01179 + C00004 + C00080	C00003 + C04367 <=> C00004 + C01244	75,05	0,00	0,977
R03350	C04478 + C00001 <=> C01187 + C00009	C04442 <=> C00022 + C00118	13,09	0,34	0,954
R03351	C00063 + C01187 <=> C00013 + C04121	C00063 + C00103 <=> C00013 + C00501	-2,35	0,69	0,892
R03378	C00044 + C16741 <=> C00035 + C03366	C00020 + C00044 <=> C00008 + C00035	-43,42	0,00	0,909
R03448	C04277 + C00001 <=> C01255	C00001 + C16512 <=> C00189 + C00249	22,52	0,00	0,814
R03470	C01278 <=> C03586 + C00011				
R03567	C02657 <=> C01467 + C00011				
R03614	C01575 + C00003 <=> C04351 + C00004 + C00080	C00003 + C00556 <=> C00004 + C00261	63,23	0,15	0,946
R03638	C01617 + C00007 + C00005 + C00080 <=> C03052 + C00006 + C00001	C00006 + C00590 <=> C00005 + C02666	56,42	0,17	0,887
R03723	C03910 <=> C01802 + C00084	C05439 <=> C05437	-6,80	0,00	0,725
R03761	C01892 + C00004 + C00080 <=> C04371 + C00003	C00003 + C02147 <=> C00004 + C00725	48,63	1,15	0,964
R03782	C02289 <=> C01951				
R03908	C02930 <=> C02295	C02930 <=> C02295	3,75	0,00	1,000
R03954	C02470 + C00001 <=> C05645				
R03959	C04105 <=> C02480	C04105 <=> C02480	50,92	1,45	1,000
R04031	C02730 + C00001 <=> C05817				
R04059	C04167 + 2 C00006 <=> C02807 + 2 C00005 + 2 C00080	C00006 <=> C00005	22,59	0,00	0,928
R04133	C04052 <=> C03063 + C00011				
R04141	C03837 + C00001 <=> C03081 + C00033				
R04151	C04468 + C00003 <=> C03164 + C00004 + C00080	C00003 + C04367 <=> C00004 + C01244	75,05	0,00	0,900
R04188	C03284 + C00026 <=> C06002 + C00025	C00026 + C00334 <=> C00025 + C00232	4,89	2,10	0,857
R04209	C04751 <=> C03373 + C00011	C04823 <=> C00122 + C04677	5,11	0,00	0,779
R04229	C03474 + 2 C00005 + 2 C00080 <=> C04050 + 2 C00006 + C00001	C00006 <=> C00005	22,59	0,00	0,933
R04259	C04522 <=> C03585				
R04260	C04105 <=> C03586	C04105 <=> C02480	50,92	1,45	0,718
R04277	C03671 + C00001 <=> C04434				

## Resultados e discussão

R04320	C00470 + C00001 <=> C00470 + C00333	C00333 <=> C00558	1,40	2,22	0,848
R04343	C00617 + C00470 <=> C00015 + C00470	C00167 <=> C00617	-1,10	0,92	0,962
R04367	C03995 <=> C02126 + C04815				
R04369	C00002 + C00133 + C03999 <=> C00008 + C00009 + C04457	C00002 + C00008 <=> C00020 + C03483	52,06	0,00	0,899
R04371	C05662 <=> C04002 + C00001	C00311 <=> C00001 + C00417	1,44	0,37	0,731
R04376	C04044 + C00007 <=> C04479				
R04378	C04677 + C00013 <=> C04051 + C00119	C00288 + C03373 <=> C00001 + C04751	5,99	0,00	0,796
R04439	C04272 + C00006 <=> C06010 + C00005 + C00080	C00006 + C00522 <=> C00005 + C00966	65,40	0,00	0,997
R04441	C04272 <=> C00141 + C00001	C03652 <=> C00022 + C00163	0,10	0,00	0,864
R04455	C04330 + C00282 <=> C04377 + C00080	C00001 + C00037 + C04377 <=> C00065 + C01217	16,01	0,00	0,913
R04520	C00353 + C04590 <=> C00013 + C04638				
R04559	C04823 <=> C00122 + C04677	C04823 <=> C00122 + C04677	5,11	0,00	1,000
R04580	C00002 + C04722 + C00010 <=> C00020 + C00013 + C04760	C00002 + C00010 + C00163 <=> C00013 + C00020 + C00100	47,12	0,00	0,951
R04602	C04798 + C00007 <=> C05155				
R04623	C04875 + C00006 <=> C04883 + C00005 + C00080	C00006 + C03351 <=> C00005 + C03067	55,18	0,00	0,912
R04809	C05447 + C00001 <=> C05448	C05268 <=> C00001 + C05271	1,03	0,00	0,962
R04812	C05450 + C00003 <=> C05467 + C00004 + C00080	C00003 + C05268 <=> C00004 + C05269	67,04	3,39	0,986
R04813	C05450 <=> C05460 + C00001	C05268 <=> C00001 + C05271	1,03	0,00	0,962
R04891	C05583 + C00003 + C00001 <=> C05584 + C00004 + C00080	C00003 + C04367 <=> C00004 + C01244	75,05	0,00	0,938
R04987	C05810 + C00007 + C00005 + C00080 <=> C05811 + C00006 + C00001	C00006 + C03351 <=> C00005 + C03067	55,18	0,00	0,965
R05078	C06033 <=> C06035 + C00001				
R05092	C06089 <=> C06090 + C00013				
R05153	C06328 + C00007 <=> C06329				
R05258	C00472 + C00088 + C00006 + C00001 <=> C00870 + C00007 + C00005 + C00080	C00006 + C03351 <=> C00005 + C03067	55,18	0,00	0,959
R05347	C07216 + C00003 <=> C07209 + C00004 + C00080	C00003 + C00556 <=> C00004 + C00261	63,23	0,15	0,997
R05356	C06593 + C00001 <=> C02378	C00001 + C16512 <=> C00189 + C00249	22,52	0,00	0,814
R05391	C03585 <=> C04431 + C01327				
R05507	C02378 + C00026 <=> C06102 + C00025	C00026 + C00956 <=> C00025 + C00322	-0,72	0,00	0,934
R05601	C06723 + C00001 <=> C06714	C05268 <=> C00001 + C05271	1,03	0,00	0,981

## Resultados e discussão

R05613	C04217 + C00129 <=> C01230 + C00013					
R05637	C11436 <=> C11453 + C00055	C00063 + C00346 <=> C00013 + C00570	-1,03	0,00	0,916	
R05639	C05108 + C00006 <=> C11455 + C00005 + C00080	C00003 + C00473 <=> C00004 + C00376	53,76	0,01	0,953	
R05697	C11500 + C00007 + C00001 <=> C11501 + C00027	C00001 + C00212 <=> C00121 + C00147	-9,77	0,00	0,871	
R05773	C07303 <=> C00755 + C00024	C03561 <=> C00001 + C00877	4,26	0,00	0,886	
R05790	C11516 + C00001 <=> C11544 + C00009	C00095 + C00118 <=> C00085 + C00577	3,31	0,00	0,757	
R05791	C10517 + C00001 <=> C11546 + C00009	C00001 + C01113 <=> C00009 + C00124	-11,55	0,00	0,888	
R06117	C11409 + C00003 <=> C11383 + C00004 + C00080	C00003 + C00756 <=> C00004 + C01545	67,46	0,00	0,955	
R06301	C11901 <=> C11878 + C00013					
R06302	C11901 <=> C11879 + C00013					
R06303	C11901 <=> C11880 + C00013					
R06378	C11397 + C00003 <=> C11398 + C00004 + C00080	C00003 + C00854 <=> C00004 + C00414	50,95	1,12	0,981	
R06380	C11398 <=> C11401					
R06382	C11399 + C00003 <=> C11401 + C00004 + C00080	C00003 + C00854 <=> C00004 + C00414	50,95	1,12	0,981	
R06383	C11413 + C00003 <=> C11415 + C00004 + C00080	C00003 + C00854 <=> C00004 + C00414	50,95	1,12	0,981	
R06417	C02485 + C00005 + C00080 <=> C11951 + C00006	C00006 <=> C00005	22,59	0,00	0,964	
R06420	C09893 + C00005 + C00080 <=> C11952 + C00006	C00006 <=> C00005	22,59	0,00	0,976	
R06532	C05906 + C00005 + C00080 <=> C06562 + C00006 + C00001	C00006 + C00590 <=> C00005 + C02666	56,42	0,17	0,908	
R06533	C05909 + C00005 + C00080 <=> C12127 + C00006 + C00001	C00006 + C00590 <=> C00005 + C02666	56,42	0,17	0,908	
R06539	C02906 + C00026 + C00007 <=> C10107 + C00042 + C00011 + C00001					
R06541	C12128 + 2 C00006 <=> C05904 + 2 C00005 + C00080	C00006 + C00590 <=> C00005 + C02666	56,42	0,17	0,866	
R06542	C09727 + 2 C00006 <=> C05905 + 2 C00005 + C00080	C00006 + C00590 <=> C00005 + C02666	56,42	0,17	0,864	
R06543	C12136 + 2 C00006 <=> C05908 + 2 C00005 + C00080	C00006 + C00590 <=> C00005 + C02666	56,42	0,17	0,864	
R06583	C05619 + C00010 + C00002 <=> C12203 + C00020 + C00013	C00002 + C00010 + C00033 <=> C00013 + C00020 + C00024	32,20	13,67	0,893	
R06615	C03648 + C00005 + C00080 <=> C09320 + C00006 + C00001	C00006 + C00590 <=> C00005 + C02666	56,42	0,17	0,910	
R06784	C11588 + C00003 <=> C04044 + C00004 + C00080	C00003 + C00473 <=> C00004 + C00376	53,76	0,01	0,928	
R06786	C11457 + C00007 + C00004 + C00080 <=> C04044 + C00001 + C00003	C00003 + C00556 <=> C00004 + C00261	63,23	0,15	0,952	
R06857	C12837 + C00003 <=> C05618 + C00004 + C00080	C00003 + C00556 <=> C00004 + C00261	63,23	0,15	0,941	

## Resultados e discussão

R06866	C05313 <=> C05802 + C00011				
R06892	C10700 + C00005 + C00080 + C00007 <=> C13636 + C00006 + C00001	C00006 + C00590 <=> C00005 + C02666	56,42	0,17	0,962
R06957	C13455 + C00001 + C00007 <=> C06082 + C00027				
R06975	C00058 + C00002 + C04677 <=> C00008 + C00009 + C04734	C00002 + C00455 <=> C00003 + C00013	-0,64	0,00	0,845
R06987	C00109 + C00010 <=> C00100 + C00058	C00024 + C00058 <=> C00010 + C00022	17,27	0,00	0,986
R06991	C14721 <=> C12247				
R07136	C11537 + C00003 <=> C05528 + C00004 + C00080	C00003 + C00258 <=> C00004 + C01146	75,66	3,71	0,930
R07137	C11537 + C00006 <=> C05528 + C00005 + C00080	C00002 + C00033 <=> C00008 + C00227	78,28	1,57	0,926
R07271	C00235 + C01263 <=> C00013 + C15509				
R07272	C00235 + C01263 <=> C00013 + C15510				
R07404	C00002 + C03373 + C00288 <=> C00008 + C00009 + C15667	C00002 + C01563 <=> C00008 + C00169	16,68	2,46	0,891
R07430	C15789 + C00007 + C00005 <=> C15790 + C00006 + C00001	C00006 <=> C00005	22,59	0,00	0,967
R07433	C10434 + C00010 <=> C00323 + C00493	C00001 + C06322 <=> C06749	-0,01	0,00	0,859
R07633	C16069 <=> C00022 + C11481				
R07649	C00448 <=> C16141 + C00013				
R07726	C10200 <=> C10205				
R07727	C16234 + 2 C00005 + 2 C00080 <=> C16235 + 2 C00006 + C00001	C00006 <=> C00005	22,59	0,00	0,944
R07736	C16188 <=> C16189				
R07788	C15797 <=> C16251				
R07801	C16263 + C00007 <=> C16264				
R07839	C04434 <=> C04324				
R07984	C04226 + 2 C00001 <=> C16159 + C00014				
R07987	C00540 + 3 C00083 <=> C16404 + 4 C00010 + 3 C00011	C01213 <=> C00683	0,00	0,00	0,871
R08058	C01217 + C00067 <=> C04377 + C00001	C00001 + C00037 + C04377 <=> C00065 + C01217	16,01	0,00	0,997
R08059	C04377 + C00006 <=> C04330 + C00005	C00006 + C00143 <=> C00005 + C00445	9,13	0,00	0,918
R08093	C16468 + C00001 <=> C16469	C03561 <=> C00001 + C00877	4,26	0,00	0,968
R08466	C08301 <=> C16719				
R08625	C17215 + C00003 <=> C17216 + C00011 + C00004 + C00080	C00001 + C00003 + C05662 <=> C00004 + C00288 + C00322	44,81	0,00	0,967
R08629	C17219 + C00003 <=> C17220 + C00011 + C00004 + C00080	C00001 + C00003 + C05662 <=> C00004 + C00288 + C00322	44,81	0,00	0,973

## Resultados e discussão

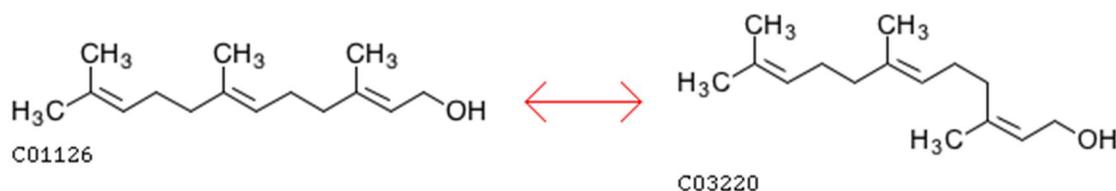
R08734	C15613 <=> C17343	2 C00024 <=> C00010 + C00332	25,68	2,29	0,951
R08766	C03672 <=> C00811 + C00001				
R08831	C05587 + C08538 <=> C17754 + C00001				
R08837	C17758 <=> C08538 + C00001				
R08842	C17733 <=> C17735				
R08853	C16269 + C00005 + C00080 + C00007 <=> C17955 + C00006 + C00001	C00006 <=> C00005	22,59	0,00	0,961
R08992	C00353 + 2 C00001 <=> C09183 + C00013				
R09043	C18132 <=> C18133	C00828 <=> C05819	-65,58	0,00	0,779
R09059	C02139 + C00001 <=> C18154				
R09111	C11901 <=> C18211 + C00013				
R09119	C06089 <=> C18226 + C00013				
R09120	C06089 <=> C11877 + C00013				
R09121	C06089 <=> C18228 + C00013				
R09214	C18302 + C00007 <=> C18303				
R09283	C06028 <=> C18323	C03561 <=> C00001 + C00877	4,26	0,00	0,951
R09292	C20238 <=> C18324	C01213 <=> C00091	-7,57	0,21	1,000
R09358	C18796 <=> C02845 + C01326				
R09487	C16143 + C00001 <=> C16286 + C00207				
R09545	C19675 + 2 C00003 + C00001 <=> C11537 + 2 C00004 + 2 C00080	C00003 + C00258 <=> C00004 + C01146	75,66	3,71	0,935
R09614	C00448 <=> C19740 + C00013				
R09664	C19791 + C00001 <=> C19792 + C00009	C00095 + C00118 <=> C00085 + C00577	3,31	0,00	0,757
R09751	3 C00088 <=> 2 C00533 + C00244 + C00001				
R09783	C19892 + C00003 + C00001 <=> C08607 + C00004 + C00080	C00003 + C00473 <=> C00004 + C00376	53,76	0,01	0,991
R09835	C19973 + C00001 <=> C19974				
R09874	C00353 + C00001 <=> C20158 + C00013				
R09967	C20270 <=> C00353 + C00001				
R10072	C16269 + 2 C00005 + 2 C00080 + 2 C00007 <=> C17954 + 2 C00006 + 3 C00001	C00006 <=> C00005	22,59	0,00	0,961
R10111	C20367 <=> C20366 + C00088				

## ***Resultados e discussão***

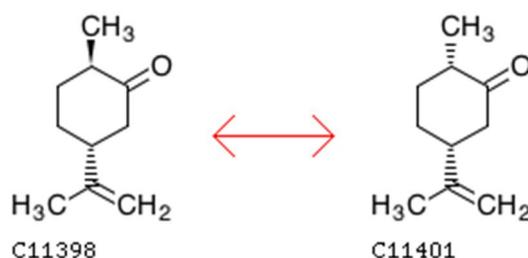
R10133	C14531 + C00001 <=> C20382 + C00088				
R10306	3 C00019 + C04441 <=> 3 C00021 + C01573	C00001 + C00021 <=> C00155 + C00212	34,74	0,00	0,812
R10350	C20525 + C00007 + C00005 + C00080 <=> C20526 + C00006 + C00001	C00006 <=> C00005	22,59	0,00	0,936
R10829	C20887 + C00009 <=> C00140 + C00636	C04256 <=> C00357	-4,60	0,00	0,874

## 3.1.1. REAÇÕES DE ISOMERIZAÇÃO

Pela **Tabela 3.1.** podemos constatar que há um total de 27 reações para as quais CC atribui um valor de 0,0, tanto para a variação na energia livre de Gibbs como para a respetiva estimativa de erro. Destas, 6 correspondem a reações de isomerização (R03265, R06380, R07839, R08734, R08842 e R09283). Quanto a estas, note-se em primeiro lugar que CC é insensível a questões de estereoquímica, e que reações como R03265 ou R06380 correspondem precisamente a transformações estereoquímicas, como ilustrado nas **Fig. 3.2.** e **Fig. 3.3.** Este é igualmente o caso das reações R08734 e R08842. Para estes e outros eventuais casos semelhantes, CC resultará sempre numa estimativa de zero para a variação na energia livre de Gibbs, mesmo em casos para o quais a transformação possa resultar em constrangimentos espaciais que impliquem que o saldo energético possa ser ligeiramente diferente de zero.



**Fig. 3.2.** – KEGG Reaction R03265. Descreve a transformação de um isómero *trans* (C01126) num isómero *cis*. Como se trata de um método insensível à estereoquímica, do ponto de vista de CC, não ocorre qualquer transformação entre reagente e produto. [Imagem adaptada de [52]]

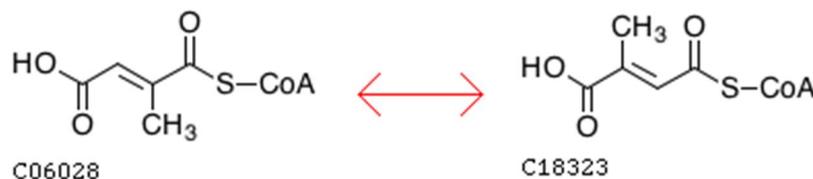


**Fig. 3.3.** – KEGG Reaction R06380. Descreve a inversão da orientação de um grupo metil (-CH<sub>3</sub>). Como se trata de um método insensível à estereoquímica, do ponto de vista de CC, não ocorre qualquer transformação entre reagente e produto. [Adaptado de [52]]

Há ainda o caso da reação R09283, ilustrada na **Fig. 3.4.**. Esta reação já não reflete uma transformação em termos estereoquímicos, mas antes a transferência de um grupo metil entre carbonos diferentes da mesma molécula. Nesta situação, estamos perante uma reação que não pode ser resolvida por RC (os seus intervenientes não podem ser

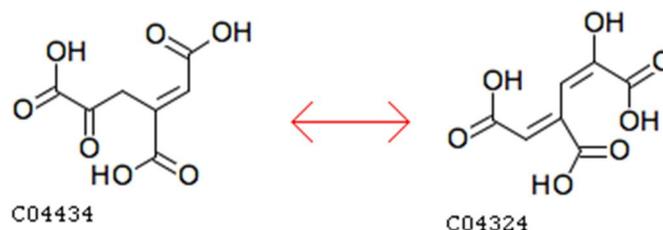
## Resultados e discussão

incluídos numa componente  $x_R$ ) e a sua decomposição em grupos solucionáveis por  $GC$  será exatamente igual de ambos os lados da equação. A estimativa gerada corresponderá assim a zero, ainda que o saldo energético da transferência deste grupo possa não ser necessariamente nulo.



**Fig. 3.4. - KEGG Reaction R09283.** Descreve a transferência de um grupo metil (-CH<sub>3</sub>) entre carbonos distintos de uma mesma molécula. Do ponto de vista de  $CC$ , casos semelhantes cujos intervenientes não possam ser resolvidos por  $RC$ , poderão resultar numa decomposição em grupos solucionáveis por  $GC$  exatamente igual entre reagentes e produtos. [Imagem adaptada de [52]]

É também importante destacar o caso de R07839, ilustrado na Fig. 3.5.. Trata-se neste caso de uma transferência de uma ligação dupla, cuja decomposição em grupos será provavelmente diferente entre o lado dos reagentes e o lado dos produtos. Contudo, segundo  $CC$  o saldo energético desta transformação será nulo.

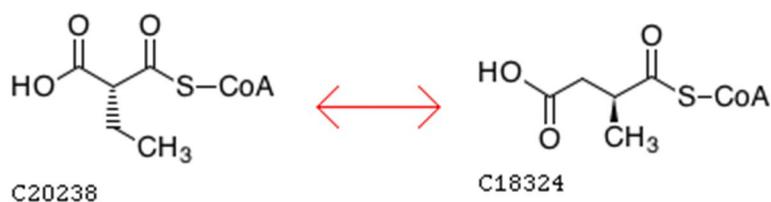


**Fig. 3.5. - KEGG Reaction R07839.** Descreve uma transferência de uma ligação dupla a partir de um grupo C=O, com a consequente formação de um grupo C=C. A estimativa obtida por  $CC$  para a variação da energia livre de Gibbs nesta reação é de zero. [Imagem adaptada de [52]]

Nesta situação concreta, há uma ligação dupla C=O que é transferida para C=C, passando o átomo de oxigénio envolvido a estar ligado a um átomo de hidrogénio num grupo -OH. Coincidentemente, e atendendo à enumeração dos grupos usados nas decomposições e às energias de formação que lhes são atribuídas<sup>[18,20]</sup>, esta situação é devidamente compensada, justificando o nulo na variação da energia de Gibbs. Mais concretamente, considera-se que do lado dos reagentes existe uma cetona (>CO,  $\Delta G_f^\circ = -115,3 \text{ kJ/mol}$ ) e um carbono secundário (-CH<sub>2</sub>-,  $\Delta G_f^\circ = 6,5 \text{ kJ/mol}$ ), que são transformados num grupo hidroxilo ligado a um carbono terciário (>COH-,  $\Delta G_f^\circ = -140,3 \text{ kJ/mol}$ ) e num carbono terciário (-CH<,  $\Delta G_f^\circ = 31,5 \text{ kJ/mol}$ ). Fazendo o somatório destas

energias de formação tanto nos reagentes como nos produtos, temos de ambos os lados um total de  $-108,8 \text{ kJ/mol}$ , pelo que a diferença entre os dois é nula. No final,  $\Delta G_r^0$  de toda a reação é zero.

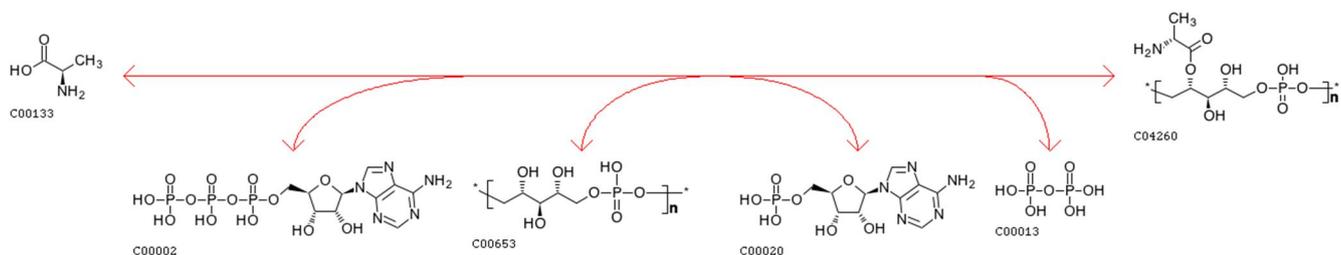
Finalmente, R09292 é igualmente um caso interessante, no qual ocorre uma mudança no esqueleto carbonado de uma molécula (Fig. 3.6.). Seria expectável que a transformação ocorrida pudesse resultar numa variação da energia livre diferente de zero, ao contrário do que acontece.



**Fig. 3.6. – KEGG Reaction R09292.** Descreve a transferência de um carbono a partir de uma cadeia lateral para a cadeia carbonada principal, resultando num grupo metil onde anteriormente se encontrava um grupo etil. Esta transformação implica ainda a deslocalização de um grupo  $-\text{COOH}$ . A estimativa obtida por *CC* para a variação da energia livre de Gibbs nesta reação é de zero. [Imagem adaptada de [52]]

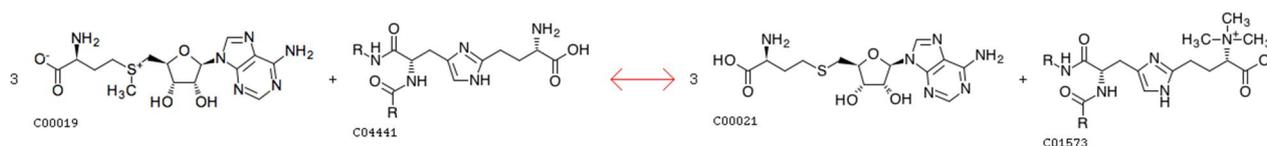
### 3.1.2. REAÇÕES QUE ENVOLVEM MOLÉCULAS COM ESTRUTURAS DESCONHECIDAS

Outro artefacto da aplicação de *CC* a qualquer conjunto de reações prende-se com a necessidade inerente a este método de se conhecer ao detalhe a estrutura química de todos os metabolitos envolvidos em cada reação. São disso exemplo as reações R02718 e R10306, ilustradas nas Fig. 3.7. e Fig. 3.8., respetivamente.



**Fig. 3.7. – KEGG Reaction R02718.** Descreve a adição de uma molécula de D-alanina a um polímero, cuja estrutura é abreviada para uma unidade repetitiva entre parêntesis e associada ao índice  $n$ . Em *CC* existe a necessidade de que a estrutura de todos os reagentes e produtos seja completamente conhecida para que se possam gerar estimativas para a variação da energia livre de Gibbs. [Imagem adaptada de [52]]

## Resultados e discussão



**Fig. 3.8. – KEGG Reaction R10306.** Descreve a transferência de 3 grupos metil (-CH<sub>3</sub>) para um aminoácido ligado a uma cadeia peptídica. A representação da cadeia peptídica é abreviada pela colocação de 2 grupos -R, que geralmente designam estruturas que podem variar. Em CC existe a necessidade de que a estrutura de todos os reagentes e produtos seja completamente conhecida para que se possam gerar estimativas para a variação da energia livre de Gibbs. [Imagem adaptada de [52]]

No caso de R02718 trata-se de uma reação que envolve um polímero. Mais concretamente, ocorre a ligação de uma molécula de D-alanina a uma cadeia polimérica, potenciada pela quebra de uma molécula de ATP. Do ponto de vista prático da aplicação de CC, é o desconhecimento do valor do índice *n* (ver Fig. 3.7.) – que designa o número de monómeros que compõem a cadeia – que impede a obtenção de uma estimativa. Neste e em casos equiparáveis, é artificialmente atribuído um valor nulo à estimativa para a variação na energia livre de Gibbs. Outros exemplos equiparáveis a este são os de R01821, R02121, R04320 e R04343.

Note-se que este caso concreto da influência de uma reação envolvendo polímeros é diferente daqueles que foram abordados na secção 2.3.. A exclusão de algumas reações com polímeros deveu-se a limitações no processamento de coeficientes estequiométricos genéricos, como os ilustrados na Fig. 2.8. (por exemplo, '*n*' ou '*n*+1'). No caso da reação R02718, há coeficientes estequiométricos definidos e o mecanismo de reação é devidamente explícito, estando a limitação da aplicação de CC relacionada com detalhes das moléculas que participam na reacção (nomeadamente, o número exato de monómeros envolvidos). Note-se ainda que estas moléculas podem ser e são devidamente representadas em ficheiros MDL, de onde constam as informações relativas ao significado do índice *n* na descrição de estruturas poliméricas.

Por sua vez, R10306 inclui a participação de moléculas cuja estrutura é representada com recurso a grupos -R (ver Fig. 3.8.). Por outras palavras, embora a reacção em si mesma esteja devidamente caracterizada, partes das estruturas dos seus reagentes e/ou produtos são de natureza variável, pelo que se opta por uma representação genérica (neste caso, a letra 'R'). Também neste caso surgem dificuldades na aplicação de CC: os grupos -R são também considerados como estruturas desconhecidas, impossibilitando a aplicação do método. À semelhança do caso anterior, é atribuído um valor nulo à estimativa gerada.

Nos casos aqui especificados, é notória a incapacidade de *CC* para gerar estimativas quando a decomposição em grupos é impossível e a aplicação de todo o algoritmo é invalidada. Assumindo que o cálculo da diferença energética entre os dois lados da equação pode ser centrado apenas no centro reativo e na sua vizinhança imediata – como, de resto, está subjacente à construção de *ReactionFingerprinter* – o problema poderia ser resolvido ignorando as estruturas desconhecidas. Por exemplo, a presença de grupos *-R* (que não interferem no mecanismo de reação) poderia ser resolvida pela sua substituição por átomos de hidrogénio, ou qualquer outro grupo com  $\Delta G_f^\theta$  conhecida. Desta forma, *CC* já seria capaz de processar a reacção, obtendo-se muito provavelmente uma estimativa num valor pouco distante da realidade. Além disto, e uma vez que a presença de grupos *-R* é especificada nos ficheiros MDL, este procedimento poderia ser incluído no próprio algoritmo do método sem dificuldades de maior e sem a necessidade de efetuar a substituição manual no ficheiro. Este tipo de procedimento poderia também ser aplicado ao caso das reacções que envolvem polímeros e em que não é especificado o número de monómeros já ligados. Uma vez que esta propriedade também integra as informações patentes nos ficheiros MDL, a presença de estruturas repetitivas que não influenciam diretamente a transformação poderia ser equiparada a uma forma de elemento neutro que não interfira com uma decomposição em grupos.

Não obstante, é também de salientar que *ReactionFingerprinter* resolveu todos os casos envolvendo grupos *-R* que *CC* não pôde processar, e ainda casos como o da reacção R02718. Este poderá ser um dos pontos fundamentais de uma eventual relação de complementaridade a aplicar entre os dois métodos.

---

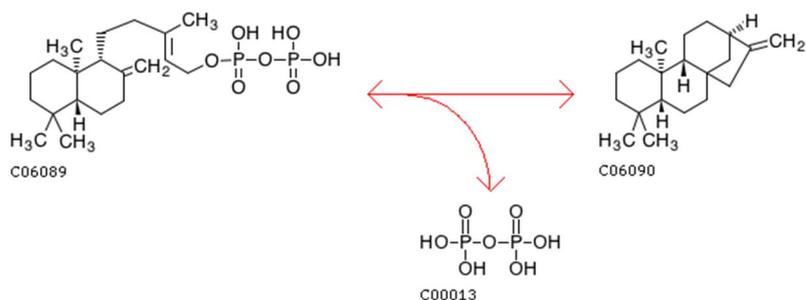
### 3.1.3. REAÇÕES QUE ENVOLVEM MOLÉCULAS QUE NÃO PODEM SER DECOMPOSTAS EM GRUPOS

---

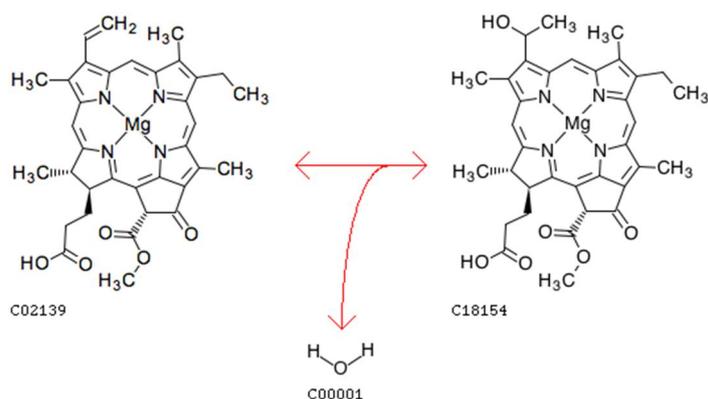
Uma outra limitação de *CC* reside em reacções nas quais participam moléculas que, não sendo solucionáveis por *RC*, também não podem ser decompostas em grupos ao abrigo de *GC*. Geralmente, estas são moléculas com grupos invulgares e que não estão previstos em *GC*. Alguns exemplos frequentes são estruturas cíclicas invulgares ou formas iónicas de ocorrência pouco comum. São os casos de R03024, R04229, R04602,

## Resultados e discussão

R05092, R05391, R07633, R07736, R08853, R09059 e R10072. As **Fig. 3.9.** e **Fig. 3.10.** ilustram, respetivamente, R05092 e R09059 – casos exemplificativos de estruturas complexas ou aberrantes que não podem ser analisadas por *GC*. Em todos os exemplos enumerados é atribuído um valor de zero à variação da energia livre de Gibbs.



**Fig. 3.9.** – KEGG Reaction R05092. Descreve a desfosforilação de copalil (C06089), com a consequente formação de Kaurene (C06090). Algumas das estruturas cíclicas de Kaurene são particularmente ilustrativas de ligações e grupos fora do comum e que não são solucionáveis pela decomposição levada a cabo por *GC*. [Imagem adaptada de [52]]



**Fig. 3.10.** – KEGG Reaction R09059. Descreve a hidroxilação de um complexo da família das clorofilas (C02139 e C18154). Estas são estruturas complexas, que ultrapassam a capacidade de decomposição em grupos proporcionada por *GC*. [Imagem adaptada de [52]]

No que diz respeito a semelhantes casos, o problema poderá ser de resolução um pouco mais complexa. Sabemos à partida que normalmente estamos perante estruturas invulgares e, por vezes, intrincadas, difíceis de prever em *GC*. Além disso, avaliar  $\Delta G^\theta$  dos grupos que eventualmente advêm da sua decomposição pode também ser uma tarefa difícil de levar a cabo. Uma sugestão para resolver pelo menos parte destes casos pode passar pela divisão destas reacções em dois sub-grupos: um em que as estruturas invulgares ou complexas intervêm diretamente na transformação ocorrida, e outro que englobe todas as situações em que, havendo uma ou mais estruturas fora do comum, as mesmas não integram o mecanismo de reação e permanecem inalteradas de ambos os lados da equação.

Evidentemente, as reações incluídas no primeiro caso serão de difícil resolução e mesmo uma transferência para serem processadas por *ReactionFingerprinter* poderá revelar-se infrutífera – se a transformação envolve estruturas marcadamente invulgares, então será pouco provável que se encontre um emparelhamento *query-target*, pelo menos dentro de um limiar de dissemelhança aceitável. Um exemplo desta situação poderá ser o da reação R05092 (ver Fig. 3.9.), na qual a transformação implica a formação de uma estrutura de anéis deveras invulgar e cuja previsão por qualquer base de dados é neste momento pouco provável.

No segundo caso podemos incluir reações como R09059 (ver Fig. 3.10.), que é um exemplo de uma transformação relativamente simples, na qual é removido um grupo -OH e formada uma nova ligação dupla =CH<sub>2</sub>. Toda a restante estrutura, apesar de volumosa e complexa, permanece inalterada, pelo que não deverá ser difícil encontrar em bases de dados como TECR pelo menos uma reacção em que ocorra uma alteração semelhante. Contudo, mesmo através de *ReactionFingerprinter* não foi possível produzir uma estimativa para este caso concreto: provavelmente, a complexidade e tamanho de uma molécula da família das clorofilas gera dissemelhanças demasiado grandes ao nível da comparação das *CF*'s de reagentes e produtos. Estas diferenças poderão ser tão grandes que absorvem por completo quaisquer similaridades evidentes nos restantes critérios utilizados para calcular um índice de semelhança. Este é outro dos pontos nos quais *ReactionFingerprinter* poderá ser aprimorado. No entanto, o facto de a documentação da *ChemAxon* não especificar a ponderação de cada critério no cálculo do índice global de semelhança será sempre um entrave no aperfeiçoamento deste aspeto.

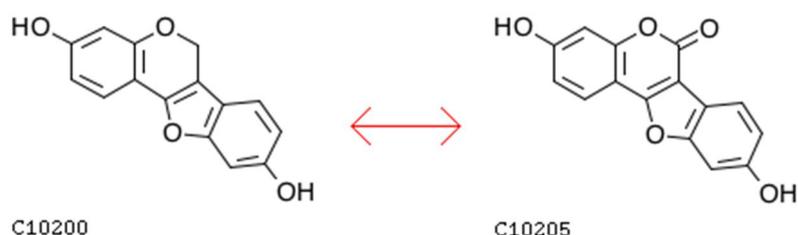
Em todo o caso, *ReactionFingerprinter* foi capaz de resolver 3 das 7 reações incluídas neste conjunto, o que significa que os dados disponíveis em TECR já englobam exemplos com os quais é possível encontrar similaridades suficientes para resolver alguns destes casos. É possível que um trabalho direcionado para a determinação experimental de  $\Delta G_r^0$  de alguns exemplos-chave de reações que envolvem estruturas invulgares permita introduzir melhorias significativas nos resultados da conjugação destes dois métodos.

3.1.4. OUTROS CASOS PARTICULARES

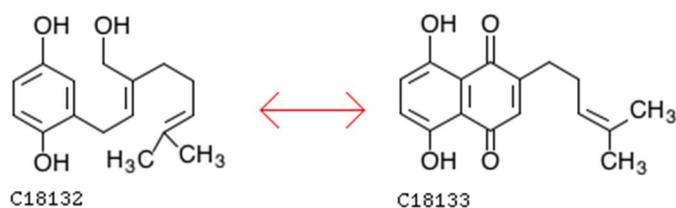
Alguns outros casos em que as estimativas para a energia livre de Gibbs são artificialmente igualadas a zero estão relacionadas com a forma como KEGG por vezes abrevia a representação de determinadas reações, nomeadamente em relação à participação de H<sub>2</sub>O, H<sup>+</sup>, ou co-fatores como NAD(P)<sup>+</sup>/NAD(P)H. Por outras palavras, há casos em que aquela base de dados omite a intervenção de moléculas que, por algum motivo, considera desnecessárias para uma representação suficientemente elucidativa da transformação ocorrida.

Do ponto de vista de CC, esta omissão é interpretada como um desequilíbrio químico. Tal como nos casos descritos nas secções anteriores, a ocorrência de um desequilíbrio químico impede a aplicação do método e as estimativas para tais reacções são, uma vez mais, igualadas a zero.

R06991, R07726, R07788, R08466 e R09043 são reações que se incluem neste grupo. As Fig. 3.11. e Fig. 3.12. ilustram, respectivamente, os exemplos de R07726 e R09043. Em ambos os casos, a análise das reações é inviabilizada pela ausência de uma molécula de H<sub>2</sub>O do lado dos reagentes, o que permitiria explicar o aparecimento de um átomo de oxigénio adicional do lado dos produtos.



**Fig. 3.11. - KEGG Reaction R07726.** Descreve uma reação na qual é adicionada ao substrato inicial uma ligação dupla a um átomo de oxigénio. A ausência de detalhes adicionais - como a possível participação de uma molécula de H<sub>2</sub>O do lado dos reagentes - leva a que a reação seja avaliada como quimicamente desequilibrada, com a consequente impossibilidade de ser analisada por CC. [Imagem adaptada de [52]]



**Fig. 3.12. - KEGG Reaction R09043.** Descreve uma reação na qual, entre outros aspetos, é adicionada ao substrato inicial uma ligação dupla a um átomo de oxigénio. A ausência de detalhes adicionais - como a possível participação de uma molécula de H<sub>2</sub>O do lado dos reagentes - leva a que a reação seja avaliada como quimicamente desequilibrada, com a consequente impossibilidade de ser analisada por CC. [Imagem adaptada de [52]]

Para estes casos, não restam muitas mais opções além da análise isolada de cada reação para que se perceba se é ou não necessário o acréscimo de algum tipo de informação. Pode ainda introduzir-se no algoritmo utilizado a capacidade de avaliar a presença de desequilíbrios químicos e procurar corrigi-los de forma automática através do acrescento das moléculas em falta. Por outro lado, esta solução pode nem sempre garantir a interpretação correta do problema, levando a que a reação que acaba por ser processada não corresponda exatamente à transformação que se pretende retratar.

Mesmo o recurso a *ReactionFingerprinter* parece não garantir a resolução deste tipo de casos, uma vez que, pelo menos na sua formulação atual, este método é insensível ao princípio de conservação da massa. Por outras palavras, *ReactionFingerprinter*, nomeadamente através do processo de *Similarity Search*, parece procurar as semelhanças entre quaisquer estruturas e/ou transformações que sejam introduzidas, independentemente da sua correção científica ou falta dela. Por exemplo, *ReactionFingerprinter* consegue processar e atribuir uma estimativa a R09043, o que leva a crer que a sua atuação não depende da correção das estruturas ou transformações analisadas. Há ainda outros casos deste género para os quais *ReactionFingerprinter* não consegue encontrar emparelhamentos dentro dos limiares definidos, e que também podem ser explicados à luz da ausência de reações suficientemente semelhantes em TEQR: assumindo que nesta base de dados as reações descritas estão devidamente formuladas e sem desequilíbrios químicos, então é improvável que se encontrem semelhanças suficientes com reações cuja transformação não é devidamente traduzida pelo respetivo ficheiro RXN.

---

### 3.1.5. REAÇÕES COM *ERR<sup>STD</sup>* MUITO ELEVADO

---

A ocorrência de um número desigual de medições – e, por vezes, a falta de qualidade de algumas delas - para as diferentes reações em TEQR conduz a algumas particularidades na forma como é estimado o erro associado a cada estimativa produzida por *CC*. Evitando entrar em demasiados detalhes, os autores de *CC*<sup>[20]</sup> atribuem um maior grau de confiança a estimativas que dependem sobretudo de reações para as quais foi realizado um grande número de medições: o grau de

## **Resultados e discussão**

conhecimento acerca da energia livre de Gibbs relativa entre dois grupos ou compostos é maior quanto maior for o número de observações experimentais sobre reações em que esses grupos ou compostos coincidam, permitindo estabelecer combinações lineares eficazes entre várias reações (atente-se no caso exemplificado na secção 1.2.2.: a coincidência do par ATP/ADP na reação **(b)** e do par H<sub>2</sub>O/P<sub>i</sub> na reação **(c)** permite estabelecer uma combinação linear que viabiliza a resolução da reação **(a)** com um elevado grau de confiança; quanto menos combinações deste género for possível fazer, menor será a confiança na estimativa calculada). Este nível de conhecimento disponível é mais tarde refletido no intervalo de confiança das estimativas obtidas - menor intervalo de confiança para casos em que as observações experimentais abundam e vice-versa.

Assumindo que as estruturas dos compostos intervenientes numa reação são bem conhecidas (por exemplo, não existem grupos *-R*) e que essas estruturas são decomponíveis em grupos previstos por *CC*, a qualidade das estimativas a produzir fica então dependente da disponibilidade de dados provenientes de observações experimentais. Essa qualidade diminui à medida que a computação de uma reação requer a consulta de dados dispersos por um grande número de reações diferentes. Quando esta dispersão é extrema, é refletida num intervalo de confiança artificial, tão grande que a estimativa gerada é completamente absorvida e perde a sua validade prática.

É este fenómeno subjacente aos valores de *Err<sup>Std</sup>* atribuídos às reações R00059, R00828, R02160, R03126, R03351, R03614, R04277, R05153, R05258, R05637, R06541, R06542, R06543, R06587, R07727, R07987, R08831, R08992, R09214, R09358, R09487, R09614, R09835, R09874, R09967, R10111 e R10133, como pode ser constatado pela leitura da **Tabela 3.1.** Este problema constitui outra das situações cuja correção dependerá maioritariamente da extensão das bases de dados disponíveis, com a adição de dados sobre novas reações, o mais variadas possível entre si.

Outra abordagem evidente seria a transferência também destes casos para *ReactionFingerprinter* ou outro método de cariz semelhante. Naturalmente que aqui será necessário fazer uma avaliação detalhada sobre se o erro associado às estimativas com base em índices de semelhança é assim tão menor do que o evidenciado por *CC*. Por um lado, *ReactionFingerprinter* depende igualmente dos dados experimentais disponíveis,

com todas as limitações que lhes estão associados: o elevado número de ensaios para alguns casos e reduzido para outros, bem como os erros tradicionalmente associados a qualquer observação experimental. Por outro, é legítimo acreditar que a propagação destes erros em métodos como *ReactionFingerprinter* será muito menor: a procura de semelhanças é sempre feita de forma independente, pelo que o erro associado à atribuição de uma estimativa dependerá do erro experimental associado ao melhor *target* e à própria assumpção de que transformações muito semelhantes produzirão um saldo energético muito próximo. (Há ainda a questão do valor do limiar de dissemelhança escolhido, mas este será sempre manipulável pelo operador, podendo ser tão aproximado de zero quanto se queira.) No final, o erro associado à estimativa para uma determinada reação será completamente independente do erro associado às estimativas para as reações anteriores e seguintes. No caso de *CC*, a propagação do erro será mais acentuada. Em primeiro lugar, os erros associados a *CC* relacionam-se também com os erros experimentais das reações de *TECR* e com a assumpção da aditividade pura das energias de formação dos grupos considerados para *GC*. Depois, note-se que  $\Delta G_f^\theta$  de um determinado composto pode ser estimado na sequência de um número virtualmente ilimitado de outras estimativas. Consideremos novamente o exemplo dado na secção 1.2.2.:  $\Delta G_r^\theta$  da reação **(a)** é estimada com base na combinação das informações disponíveis sobre as reações **(b)** e **(c)**, o que constitui em si mesmo uma forma de aproximação – ainda que este efeito possa ser quase negligenciável. Em *CC* a reacção **(a)** poderia a jusante ser utilizada da mesma forma que **(b)** e **(c)** foram aqui utilizadas, de modo a gerar uma estimativa para uma hipotética reacção **(d)**, e assim sucessivamente. Na verdade,  $\Delta G_r^\theta$  de várias reações e  $\Delta G_f^\theta$  de vários compostos foram desta forma obtidos pelos autores de *CC*.<sup>[20]</sup> Além disto, há ainda casos de  $\Delta G_f^\theta$  dos substratos de determinadas reações, que deviam preferencialmente ser estimados com base em reações nas quais ambos coexistam, entre reagentes e produtos (à semelhança daquilo que é feito com ATP e ADP na reacção **(a)**). Todavia, a indisponibilidade de dados experimentais que permitam tal exercício obriga a que em alguns casos a estimativa seja obtida por vias menos rigorosas. É especialmente nestes casos que o intervalo de confiança é amplamente alargado e a estimativa perde grande parte da sua exatidão. Esta propagação do erro entre estimativas poderá estar, pelo menos parcialmente, na origem da grande discrepância verificada entre algumas estimativas de *CC* e as

## **Resultados e discussão**

estimativas correspondentes em *ReactionFingerprinter*, com as estimativas de *CC* a apresentarem a tendência para uma maior dispersão em torno de zero (ver secção 3.2.).

Não obstante, na formulação de *CC* observa-se um grande cuidado em manter uma importante coerência entre os pontos de referência tidos por base para o cálculo de estimativas. Esta atenção constitui uma vantagem essencial deste método quando se avança para a comparação entre as estimativas produzidas para diferentes reações e especialmente relevante na análise de vias metabólicas completas. Este é um aspeto em que *ReactionFingerprinter*, pelo menos na formulação atual, está ainda em desvantagem: a atribuição de uma estimativa resulta de uma média aritmética simples entre os valores patentes em *TECR* para os vários ensaios sobre uma mesma reação, o que está longe do rigor proporcionado pelo processo de normalização seguido em *CC*.

---

### 3.1.6. EFEITO DA DEFINIÇÃO DO LIMIAR DE DISSEMELHANÇA EM *REACTION FINGERPRINTER*

---

De acordo com a **Tabela 3.2.**, os resultados conseguidos por *ReactionFingerprinter* para um limiar de 0,2 indicam que a cobertura isolada deste método é da ordem dos 58%, registando-se apenas uma pequena subida para os 63% quando se passa a considerar as estimativas geradas para limiares de 0,25 e 0,3. Sobra então algum espaço para especular sobre qual seria a melhoria obtida ao alargar o limiar de dissemelhança a valores mais elevados. Por exemplo, *IGERS* obteve resultados com margens de erro inferiores às de *GC* para limiares de 0,4.<sup>[17]</sup> Contudo, é possível que a variedade de transformações caracterizadas pelas reações de *TECR* não seja suficiente para que um alargamento do limiar de dissemelhança produza muitos mais emparelhamentos *query-target*. É, pois, legítimo supor que novos alargamentos não signifiquem um aumento considerável na cobertura proporcionada por *ReactionFingerprinter*.

À semelhança do que já foi discutido acerca de *CC*, também *ReactionFingerprinter* beneficiaria muito do alargamento das bases de dados disponíveis com informação sobre a termodinâmica de reações bioquímicas. Novamente, é possível que o acréscimo de informação sobre alguns grupos-chave de reações proporcionasse uma melhoria significativa nos resultados obtidos. Por exemplo, a análise experimental de algumas reações, devidamente selecionadas, que envolvem estruturas fora do comum (por

exemplo, os casos de R05092 e R09059, ilustrados nas Fig. 3.9. e Fig. 3.10.) poderia ser suficiente para resolver alguns casos particulares para os quais *ReactionFingerprinter* não encontrou emparelhamentos. Outro exemplo diz respeito ao estudo muito completo de um grupo numeroso de reações originárias de organismos termófilos e hipertermófilos descrito em [62]. Todos estes são casos que permitiriam acrescentar variedade a bases de dados como TECR. Tal variedade não só poderia aumentar o potencial de emparelhamento em *ReactionFingerprinter*, como ainda poderia trazer informação adicional a ser incluída em CC, quer do ponto de vista de RC como de GC.

### 3.1.7. EMPARELHAMENTO DE REAÇÕES EM SENTIDOS OPOSTOS

Outro dado curioso sobre o funcionamento de *ReactionFingerprinter* está relacionado com a busca de similaridades em reações no sentido inverso. Por exemplo, atente-se no caso da reacção R08987, ilustrado na Fig. 3.13.:

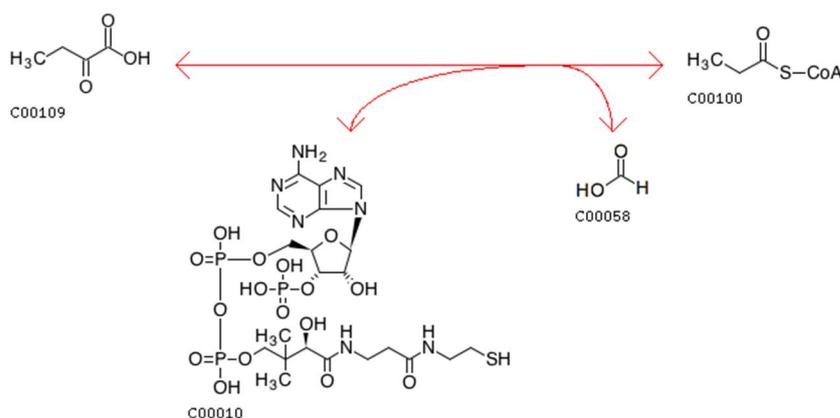
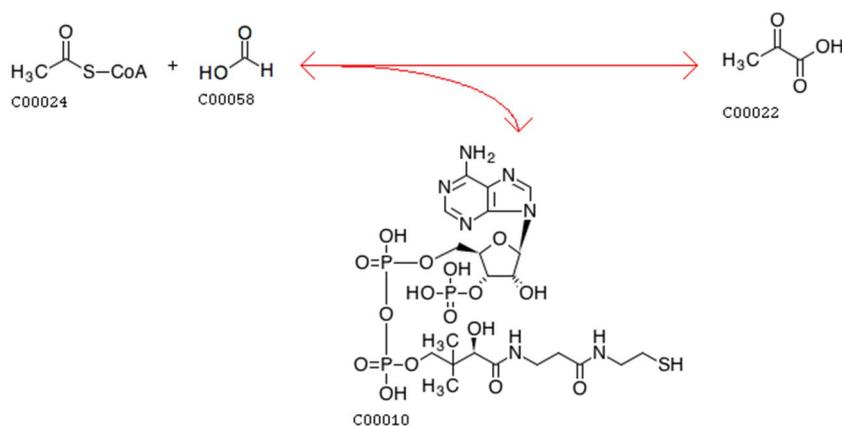


Fig. 3.13 - KEGG Reaction R06987. Descreve a transferência de um grupo propanoil para uma molécula de coenzima A, obtendo-se a formação de propanoil-CoA e formato. [Imagem adaptada de [52]]

A reacção R06987 é emparelhada por *ReactionFingerprinter* com o *target* de TECR ilustrado na Fig. 3.14., que em KEGG tem correspondência com a reacção R00212:

## Resultados e discussão



**Fig. 3.14.** – KEGG Reaction R00212. Descreve um mecanismo de quebra de Acetil-CoA, com formação de coenzima A e piruvato. Note-se que a quebra da molécula de coenzima A ocorre no sentido oposto ao da reação R06987, ilustrado na Fig. 3.12.. [Imagem adaptada de [52]]

Como se pode ver, o emparelhamento de R06987 é feito com um *target* onde a transformação analisada (formação ou quebra de uma molécula de coenzima A) ocorre no sentido inverso, indicando que a *Similarity Search* de *JChem*® é capaz de procurar semelhanças simultaneamente em ambos os sentidos. Contudo, a atribuição de um valor à energia livre de Gibbs é feita simplesmente de acordo com o que está tabelado em TEQR para o *target* emparelhado. Note-se que nestes casos, ao haver um emparelhamento com uma reação com similaridades no sentido inverso, também deveria ser atribuída uma variação da energia livre com o valor simétrico ao tabelado. Tal não acontece, com a atribuição de uma estimativa para a variação na energia livre de Gibbs de cerca de 17,27 *kJ/mol*. Pela lógica do sentido das reações exemplificadas, a estimativa atribuída deveria corresponder ao valor simétrico de -17,27 *kJ/mol*. Em todos os casos semelhantes, este será certamente um dos principais fatores a introduzir diferenças significativas entre o valor da estimativa gerada por *ReactionFingerprinter* e aquele que é obtido por *CC*.

### 3.2. AVALIAÇÃO DA COMPLEMENTARIDADE ENTRE *COMPONENT CONTRIBUTION* E *REACTION FINGERPRINTER*

Além da abrangência proporcionada por cada método isoladamente, é também importante considerar quão vasta poderá ser a cobertura proporcionada pelos dois

métodos quando avaliados em conjunto. Por exemplo, é desde logo importante levar a cabo o exercício de verificar se os casos de insucesso de um são também casos de insucesso do outro. A Tabela 3.3. elabora um quadro-resumo das reações em que cada um dos métodos revelou algum tipo de limitação, assinalando ainda os casos de insucesso comuns a ambos.

A Tabela 3.3. abrange 96 reações para as quais pelo menos um dos métodos não conseguiu gerar uma estimativa válida. Nos casos de *ReactionFingerprinter*, como já vimos, a incapacidade deve-se à inexistência de reações em TECR com as quais seja possível efectuar um emparelhamento *query-target*, logo impossibilitando a atribuição de um valor de  $\Delta G_r^0$ . Nos casos relativos às limitações de *CC*, é indicado o motivo específico pelo qual a obtenção de uma estimativa foi impossibilitada ou invalidada (estruturas desconhecidas, estruturas não decomponíveis, outros casos de não processamento e incertezas elevadas).

**Tabela 3.3.** – Quadro-resumo das reações para as quais não foi possível gerar estimativas por algum dos métodos. O 'X' indica a ausência de um resultado pelo método designado no topo da respetiva coluna – *CC* ou *ReactionFingerprinter*. (No caso de *CC*, é ainda especificada a limitação concreta que impediu o cálculo de uma estimativa.) Estão também assinalados a verde os casos em que, sendo impossível o cálculo por algum dos métodos, a cobertura é assegurada pelo outro; aqueles que não podem ser resolvidos por nenhum dos métodos estão assinalados a vermelho.

KEGG Reaction	Component Contribution				Reaction Fingerprinter	KEGG Reaction	Component Contribution				Reaction Fingerprinter
	Estrutura desconhecida	Não decomponível	Outros casos	Elevadas incertezas			Estrutura desconhecida	Não decomponível	Outros casos	Elevadas incertezas	
R00043					X	R06302					X
R00059				X		R06303					X
R00348					X	R06380					X
R00828				X		R06539				X	X
R00913					X	R06541				X	
R01085					X	R06542				X	
R01334					X	R06543				X	
R01635					X	R06587				X	
R01821	X					R06866					X
R02121	X					R06957					X
R02160				X	X	R06991			X		X
R02340					X	R07271					X
R02540					X	R07272					X
R02613					X	R07633			X		X

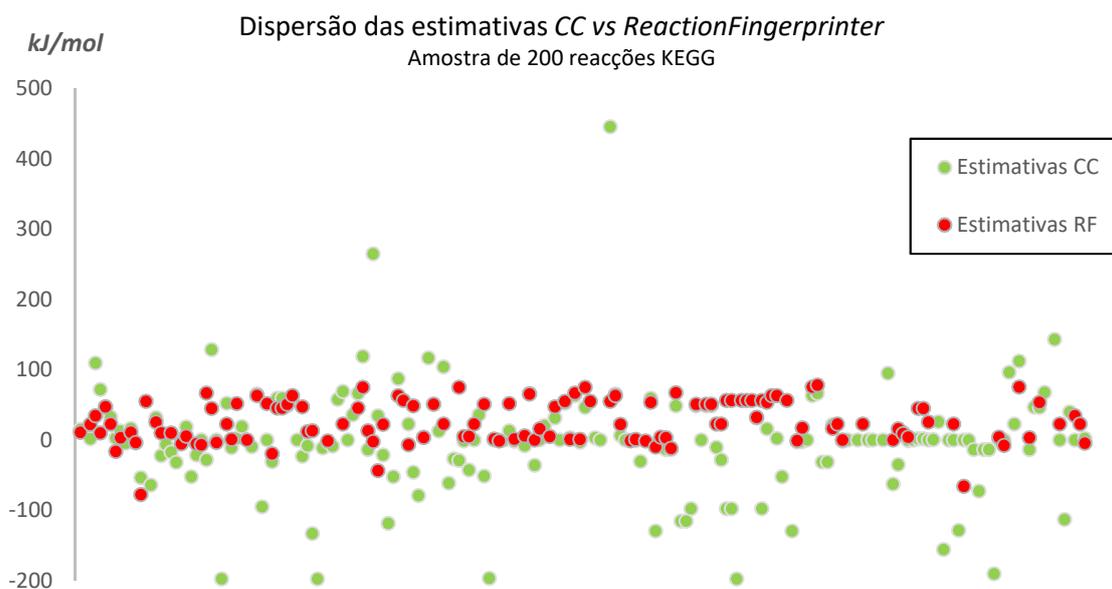
## Resultados e discussão

R02718	X			R07649			X	
R03024		X		X	R07726		X	X
R03126			X	X	R07727		X	
R03180				X	R07736		X	X
R03249				X	R07788		X	X
R03250				X	R07801			X
R03265				X	R07839			X
R03307				X	R07984			X
R03351			X		R07987		X	
R03470				X	R08466		X	X
R03567				X	R08766			X
R03614			X		R08831		X	X
R03782				X	R08837			X
R03954				X	R08842			X
R04031				X	R08853		X	
R04133				X	R08992		X	X
R04141				X	R09043		X	
R04229			X		R09059		X	X
R04259				X	R09111			X
R04277			X	X	R09119			X
R04320				X	R09120			X
R04343				X	R09121			X
R04367				X	R09214		X	X
R04376				X	R09358		X	X
R04520				X	R09487		X	X
R04602			X	X	R09614		X	X
R05078				X	R09751			X
R05092			X	X	R09835		X	X
R05153			X	X	R09874		X	X
R05258			X		R09967		X	X
R05391			X	X	R10072		X	
R05613				X	R10111		X	X
R05637			X		R10133		X	X
R06301				X	R10306		X	

Verifica-se que destas 96 reações há 27 para as quais nenhum método pôde produzir uma estimativa ('X' em mais do que uma coluna); há ainda 22 que, não podendo ser resolvidas por *CC*, puderam ser processadas por *ReactionFingerprinter* com a obtenção de uma estimativa válida. Das primeiras, constata-se que a maioria, 16, são reacções para as quais *CC* não pode produzir uma estimativa com um intervalo de confiança aceitável; 7 correspondem a casos que envolvem moléculas não

decomponíveis e 4 a reações com outro tipo de artefacto, tais como desequilíbrios químicos. Das restantes, a maioria, 12, diz também respeito a cálculos de *CC* com intervalos de confiança demasiado largos; 6 correspondem a reações que envolvem compostos com estruturas desconhecidas (grupos *-R*, estruturas poliméricas), 3 envolvem estruturas não decomponíveis em grupos e 1 contém outro tipo de problema (por exemplo, desequilíbrio químico). É ainda interessante reparar que *ReactionFingerprinter* conseguiu encontrar um emparelhamento, e, conseqüentemente, uma estimativa, para todas as 6 reações cujo problema em *CC* estava relacionado com as estruturas por definir. *ReactionFingerprinter* resolveu ainda 3 casos problemáticos em *CC* de entre os 10 devidos a estruturas não decomponíveis em grupos. Se se considerar que *ReactionFingerprinter* resolveu com sucesso 22 das 49 reações para as quais *CC* mostrou limitações, então a eficácia combinada de ambos os métodos sobe para os 86,5%, relativamente aos cerca de 74,5% obtidos com a utilização isolada de *CC*. A aplicação conjugada dos dois métodos pode assim significar um incremento de 12% na capacidade de cobertura das reações da amostra considerada.

É também importante reparar na comparação da dispersão dos valores das estimativas obtidas por cada um dos métodos, retratada no gráfico da Fig. 3.15.. Apesar de *ReactionFingerprinter* ter gerado estimativas para um menor número de reações, o que se reflecte num menor número de pontos no gráfico, parece haver uma tendência clara para que os seus valores se concentrem numa faixa mais estreita, distribuída sobretudo entre 0 e 100 *kJ/mol*.



**Fig. 3.15.** – Dispersão das estimativas geradas por CC (a verde) e por *ReactionFingerprinter* (a vermelho). As estimativas geradas por *ReactionFingerprinter* parecem localizar-se regra-geral dentro de um intervalo mais estreito.

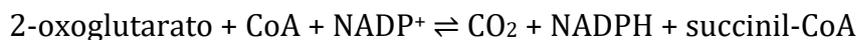
De acordo com as observações feitas em 3.1.5. e 3.1.7., tanto a forma como os erros são propagados em cada método como o cálculo de estimativas a partir de reacções no sentido inverso podem ajudar a explicar a dispersão dos dados aqui observada.

### 3.3. DESEMPENHO DE REACTION FINGERPRINTER PARA REACÇÕES COM $\Delta G_R^0$ CONHECIDO MAS AUSENTES DE TECR

Perceber qual a proximidade (ou distância) das estimativas de *ReactionFingerprinter* face à realidade envolveu a escolha de 10 reacções que não constam da base de dados TECR (mas para as quais existe um valor de  $\Delta G_R^0$  determinado por trabalhos experimentais) e analisá-las à luz deste método (ver secção 2.4.). Os resultados obtidos são retratados na Tabela 3.4., onde é possível constatar que apenas em dois destes casos foi possível obter uma estimativa. Em ambos os casos as estimativas foram obtidas com base no limiar de dissemelhança de valor mínimo: 0,2.

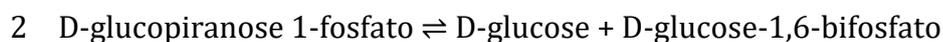
A primeira das reacções para as quais foram obtidas estimativas por *ReactionFingerprinter* advém de um trabalho no sentido de publicar uma base de dados

inovadora e consistente com informações sobre reações do ciclo dos ácidos tricarboxílicos, glicólise e via das pentoses fosfato<sup>[59, 60]</sup>:



Esta reação não consta de TECR e foi caracterizada pelos seus autores com uma energia livre de Gibbs padrão de  $15,28 \text{ kJ/mol}$ . Neste trabalho, e com base apenas em *ReactionFingerprinter*, o mesmo parâmetro foi estimado em cerca de  $60,39 \text{ kJ/mol}$ , o que corresponde a cerca de 4 vezes o valor determinado experimentalmente.

A segunda reação provém de uma publicação na qual é abordado o equilíbrio químico e a cinética enzimática durante a actividade de células musculares<sup>[61]</sup>:



Esta reação também não consta de TECR e o cálculo da sua energia livre de Gibbs por métodos experimentais cifrou-se em  $-6,67 \text{ kJ/mol}$ . Por *ReactionFingerprinter* obteve-se uma estimativa de  $-6,46 \text{ kJ/mol}$ , embora esta análise seja feita tendo por base valores provenientes de TECR, que estão normalizados para condições padrão. Sabendo que a medição original foi feita para uma temperatura de  $303,15 \text{ K}$ ,  $\text{pH} = 7,4$  e  $I = 0,1 \text{ M}$  (ver **Tabela 2.1.**), é provável que o desfasamento real entre os valores seja ligeiramente maior. Não obstante, neste exemplo a distância entre valor experimental e valor estimado será provavelmente muito mais curta do que no primeiro caso.

Infelizmente, para os restantes oito casos propostos para avaliação por *ReactionFingerprinter* não foi possível obter estimativas de acordo com os valores definidos para o limiar de dissemelhança. Convém lembrar que este é um conjunto de reações originárias do metabolismo de organismos termófilos e hipertermófilos<sup>[62]</sup>, caracterizados por uma série de particularidades fisiológicas. Tais particularidades incluem a capacidade de metabolizar substratos invulgares (como é o caso de substâncias sulfuradas e vários hidrocarbonetos), nomeadamente através de vias ausentes na maioria dos organismos habitualmente estudados e de onde provém a maior parte das reacções patentes em TECR. Como sugerido em 3.1.6., a adição de dados deste tipo de reações a bases de dados como TECR permitiria um forte alargamento à capacidade de cobertura de *ReactionFingerprinter*.

## Resultados e discussão

**Tabela 3.4.** - Estimativas para a energia livre de Gibbs,  $\Delta G_r^0$ , obtidas por *ReactionFingerprinter* para um conjunto de 10 reações para as quais a energia livre foi obtida pela via experimental, mas que não constam de TECR. As estimativas são obtidas pela atribuição do valor médio das energias livres de Gibbs disponíveis em TECR para a reação com maior índice de semelhança relativamente à reacção de *query*. A cada estimativa está ainda associado o desvio-padrão, *StdDev*, da medição e o maior índice de semelhança para o qual foi possível obter um emparelhamento *query-target*.

RF ID	KEGG Reaction	Melhor target	$\Delta_r G$ (kJ/mol)	$\Delta_r G^0$ (kJ/mol)	StdDev	Índice de semelhança
59.1	C00026 + C00010 + C00006 <=> C00011 + C00005 + C00091	C00003 + C01144 <=> C00004 + C00332	15.28	60.3930435	1,36295668	0,961
61.1	2 C00103 <=> C00031 + C01231	C00103 <=> C00092	-6.67	-6.46323464	2,205547709	1,000
62.1	4 C91008 + 2 C00080 <=> 3 C91009 + 2 C00283					
62.2	4 C00058 <=> C01438 + 3 C00011 + 2 C00001					
62.3	C00186 + 2 C91010 <=> C00033 + 2 C91011 + C00011 + C00001					
62.4	2 C00246 + 5 C91009 + 10 C00080 <=> 5 C00283 + 8 C00011 + 8 C00001					
62.5	2 C00469 + C91009 + 2 C00080 <=> C00283 + 2 C00033 + 2 C00001					
62.6	2 C05979 + C91009 + 2 C00080 <=> 2 C00163 + C00283 + 2 C00001					
62.7	C00033 + 8 C14819 + 2 C00001 <=> 8 C14818 + 2 C00011 + 8 C00080					
62.8	C00160 <=> C91012 + C00080					

---

### 3.4. ANÁLISE DE VIAS METABÓLICAS

---

A análise das vias metabólicas por *CC* e *ReactionFingerprinter* foi feita através da inserção das reações de cada via em ambos métodos, na sua ordem sequencial, seguida do cálculo do total das energias livres de Gibbs de cada reação isolada. *CC* foi capaz de gerar estimativas para todas as reações de cada via, com a exceção da primeira reação da biossíntese de violaceína, para a qual o intervalo de confiança obtido é demasiado grande. Por seu turno, *ReactionFingerprinter* voltou a revelar mais limitações, sendo que apenas para a degradação do glicerol foi capaz de gerar estimativas para todas as reações; nas restantes vias não foi possível estabelecer um contínuo de estimativas que perfaça uma estimativa final para a energia livre de Gibbs envolvida em todo o processo. Note-se que a todas as reações de cada via foi ainda conferido um '*ReactionFingerprinter ID*', no intuito de numerar cada reação numa ordem sequencial na via em que se insere.

---

#### 3.4.1. BIOSÍNTESE DE L-FENILALANINA

---

Através de *CC* foi possível traçar uma estimativa total para a biossíntese de L-fenilalanina, nomeadamente pela soma das energias livres de Gibbs de cada uma das dez reações que constituem esta via metabólica. A **Tabela 3.5.** resume as informações mais importantes de cada reação e as respetivas estimativas parciais e total.

## Resultados e discussão

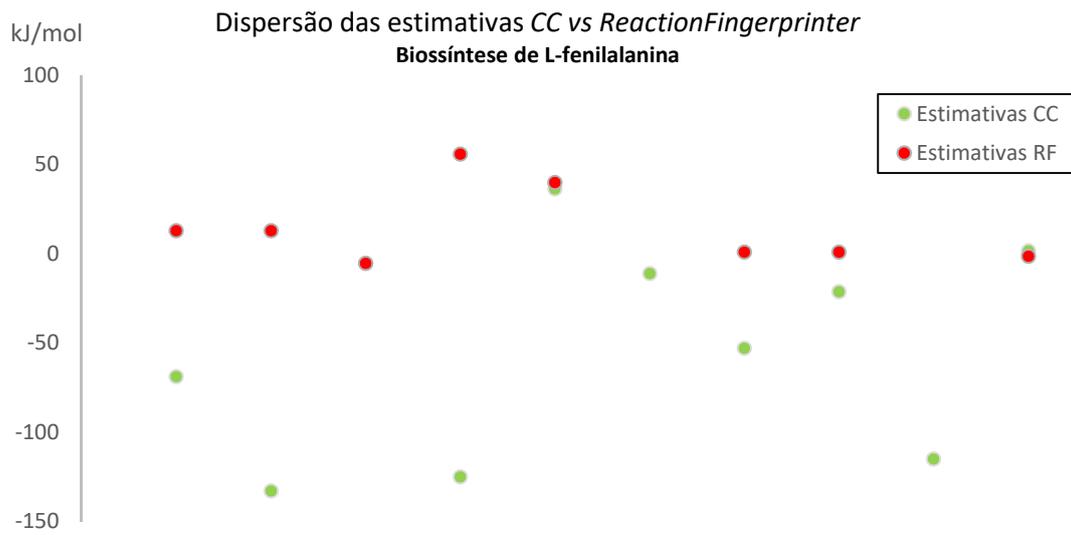
**Tabela 3.5.** – Estimativas para a energia livre de Gibbs,  $\Delta G_r^0$ , das reações da via de biossíntese de L-fenilalanina obtidas por *CC*, bem como para a energia total envolvida na totalidade da via. *Component Contribution* calcula ainda o erro padrão,  $Err^{Std}$ , associado a cada estimativa.

ReactionFingerprinter ID	KEGG Reaction	$\Delta_r G^0$ (kJ/mol)	$Err^{Std}$ (+/- kJ/mol)
PWY-6628_01_R01826	C00074 + C00279 + C00001 <=> C04691 + C00009	-68.7	4.7
PWY-6628_02_R03083	C04691 <=> C00944 + C00009	-132.8	9.8
PWY-6628_03_R03084	C00944 <=> C02637 + C00001	-5.3	4.1
PWY-6628_04_R02413	C02637 + C00005 + C00080 <=> C00493 + C00006	-124.9	7.9
PWY-6628_05_R02412	C00002 + C00493 <=> C00008 + C03175	36.5	3.3
PWY-6628_06_R03460	C00074 + C03175 <=> C00009 + C01269	-10.9	4.5
PWY-6628_07_R01714	C01269 <=> C00251 + C00009	-52.8	7.1
PWY-6628_08_R01715	C00251 <=> C00254	-21.1	8.9
PWY-6628_09_R01373	C00254 <=> C00166 + C00001 + C00011	-114.9	18.3
PWY-6628_10_R00694	C00166 + C00025 <=> C00079 + C00026	1.7	4.1
<b>Total</b>		-493.2 kJ/mol	

Pela **Tabela 3.5.** é possível observar que *CC* foi capaz de processar todas as reações envolvidas, obtendo estimativas válidas para todas. O valor total do aporte energético estimado para esta via metabólica é de -493,2 kJ/mol. De acordo com esta estimativa, trata-se de uma via claramente exergônica.

Por outro lado, *ReactionFingerprinter* foi incapaz de obter uma estimativa para duas das reações envolvidas e a partir de limiares com o valor de 0,2, 0,25 e 0,3. Em concreto, a energia livre de Gibbs da sexta e da nona reações desta sequência ficaram por estimar por falta de um emparelhamento *query-target*. A impossibilidade de obter estimativas para todas as reações impede ainda o cálculo do cômputo energético total desta via metabólica por *ReactionFingerprinter*. A **Tabela 3.6.** resume os dados obtidos para a biossíntese da L-fenilalanina por *ReactionFingerprinter*.

O gráfico da **Fig. 3.16.** ilustra de forma mais visual a diferença entre as estimativas geradas por *CC* e as obtidas por *ReactionFingerprinter*. À semelhança do observado para a amostra de 200 reações de KEGG, os valores obtidos por *ReactionFingerprinter* parecem situar-se num intervalo mais estreito do que os estimados por *CC*, e com tendência a apresentar valores mais positivos.



**Fig. 3.16.** – Dispersão das estimativas geradas por *CC* e por *ReactionFingerprinter* para as reações da via de biossíntese de L-fenilalanina. As estimativas geradas por *ReactionFingerprinter* (a vermelho) parecem localizar-se regra-geral dentro de um intervalo mais estreito e em valores mais positivos do que as estimativas geradas por *CC* (a verde).

## Resultados e discussão

**Tabela 3.6.** - Estimativas para a energia livre de Gibbs,  $\Delta G_r^0$ , obtidas por *ReactionFingerprinter* para as reações que constituem a via metabólica da biossíntese da L-fenilalanina. A cada estimativa está também associado o desvio-padrão, *StdDev*, da medição e o valor do maior índice de semelhança com o qual foi obtida.

RF ID	KEGG Reaction	Melhor target	$\Delta_r G^0$ (kJ/mol)	StdDev	Índice de semelhança
PWY-6628_01_R01826	C00074 + C00279 + C00001 $\Leftrightarrow$ C04691 + C00009	C04442 $\Leftrightarrow$ C00022 + C00118	13.09	0,34	0,822
PWY-6628_02_R03083	C04691 $\Leftrightarrow$ C00944 + C00009	C04442 $\Leftrightarrow$ C00022 + C00118	13.09	0,34	0,822
PWY-6628_03_R03084	C00944 $\Leftrightarrow$ C02637 + C00001	C00944 $\Leftrightarrow$ C00001 + C02637	-5.29	1,51	1,000
PWY-6628_04_R02413	C02637 + C00005 + C00080 $\Leftrightarrow$ C00493 + C00006	C00006 + C00493 $\Leftrightarrow$ C00005 + C02652	55.98	0,50	0,988
PWY-6628_05_R02412	C00002 + C00493 $\Leftrightarrow$ C00008 + C03175	C00002 + C00022 + C00288 $\Leftrightarrow$ C00008 + C00009 + C00036	40.12	2,74	0,872
PWY-6628_06_R03460	C00074 + C03175 $\Leftrightarrow$ C00009 + C01269				
PWY-6628_07_R01714	C01269 $\Leftrightarrow$ C00251 + C00009	C00251 $\Leftrightarrow$ C00885	0.98	0,43	0,740
PWY-6628_08_R01715	C00251 $\Leftrightarrow$ C00254	C00251 $\Leftrightarrow$ C00885	0.98	0,43	0,756
PWY-6628_09_R01373	C00254 $\Leftrightarrow$ C00166 + C00001 + C00011				
PWY-6628_10_R00694	C00166 + C00025 $\Leftrightarrow$ C00079 + C00026	C00026 + C00082 $\Leftrightarrow$ C00025 + C01179	-1.20	0,04	0,905

---

### 3.4.2. DEGRADAÇÃO DE GLICEROL E PRODUÇÃO DE BUTANOL

---

A via de degradação do glicerol com produção de butanol é constituída por quinze reações sequenciais. Contudo, as últimas seis reações pressupõem a disponibilidade de acetil-CoA no dobro da quantidade estequiométrica daquela que é produzida ao cabo das primeiras nove reações (ver secção 2.5.2.). Por este motivo, qualquer estimativa válida para o aporte energético total da via terá de levar em conta as primeiras nove reacções em duplicado, correspondendo assim ao verdadeiro custo energético da degradação do glicerol com produção de butanol.

A **Tabela 3.7.** indica as energias livres de Gibbs parciais e total desta via, estimadas através de *CC*. Novamente, *CC* apresentou-se capaz de processar todas as reações e de produzir estimativas com intervalos de confiança razoáveis. O cômputo energético total estimado é de  $-74,4 \text{ kJ/mol}$ , indiciando que se trata de uma via exergónica.

*ReactionFingerprinter* foi igualmente capaz de processar todas as reações que compõem a via de degradação do glicerol, encontrando emparelhamentos *query-target* para todos os casos. Destaca-se o facto de várias das estimativas geradas serem muito próximas das que foram obtidas por *CC*: as energias livres da primeira, quarta, sétima, décima e décima segunda reacções podem considerar-se praticamente iguais às estimadas por *CC*; embora em menor grau, também as estimativas para a quinta, nona e décima primeira reações se encontram muito próximas das que são avançadas por *CC*. Contudo as restantes sete reações apresentam números mais divergentes relativamente aos de *CC*, levando a que o aporte energético total seja estimado por *ReactionFingerprinter* em  $599,65 \text{ kJ/mol}$ , valor indicativo de uma via claramente endergónica. A **Tabela 3.8.** retrata todas estas considerações.

## Resultados e discussão

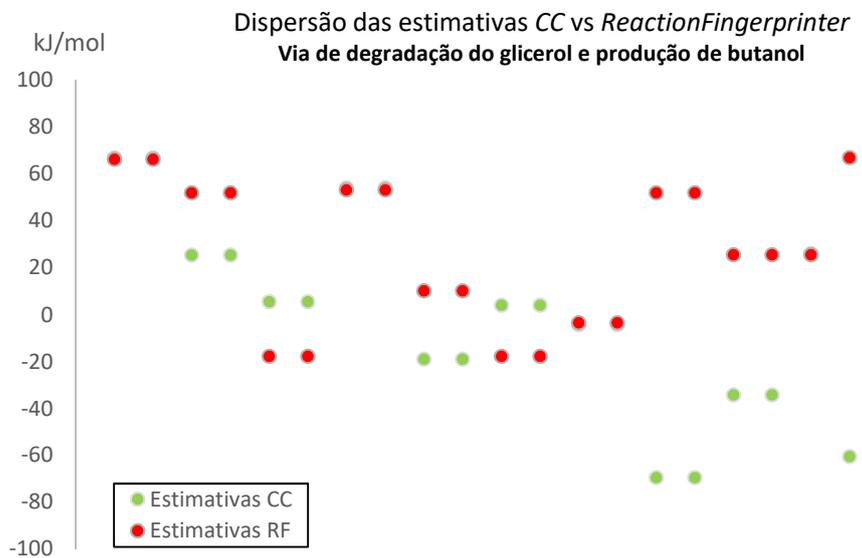
**Tabela 3.7.** - Estimativas para a energia livre de Gibbs,  $\Delta G_r^0$ , das reações da via de degradação do glicerol obtidas por CC, bem como para a energia total envolvida na totalidade da via. *Component Contribution* calcula ainda o erro padrão,  $Err^{Std}$ , associado a cada estimativa.

RF ID	KEGG Reaction	$\Delta_r G^0$ (kJ/mol)	$Err^{Std}$ (+/- kJ/mol)
PWY-7003_01.1_R01034	C00116 + C00003 <=> C00184 + C00004 + C00080	67.0	3.6
PWY-7003_01.2_R01034	C00116 + C00003 <=> C00184 + C00004 + C00080	67.0	3.6
PWY-7003_02.1_R01011	C00002 + C00184 <=> C00008 + C00111	25.5	4.3
PWY-7003_02.2_R01011	C00002 + C00184 <=> C00008 + C00111	25.5	4.3
PWY-7003_03.1_R01015	C00111 <=> C00118	5.6	1.1
PWY-7003_03.2_R01015	C00111 <=> C00118	5.6	1.1
PWY-7003_04.1_R01061	C00118 + C00009 + C00003 <=> C00236 + C00004 + C00080	54.3	0.8
PWY-7003_04.2_R01061	C00118 + C00009 + C00003 <=> C00236 + C00004 + C00080	54.3	0.8
PWY-7003_05.1_R01512	C00008 + C00236 <=> C00002 + C00197	-18.8	0.9
PWY-7003_05.2_R01512	C00008 + C00236 <=> C00002 + C00197	-18.8	0.9
PWY-7003_06.1_R01518	C00197 <=> C00631	4.1	0.7
PWY-7003_06.2_R01518	C00197 <=> C00631	4.1	0.7
PWY-7003_07.1_R00658	C00631 <=> C00074 + C00001	-4.1	0.6
PWY-7003_07.2_R00658	C00631 <=> C00074 + C00001	-4.1	0.6
PWY-7003_08.1_R00200	C00008 + C00074 <=> C00002 + C00022	-69.3	0.8
PWY-7003_08.2_R00200	C00008 + C00074 <=> C00002 + C00022	-69.3	0.8
PWY-7003_09.1_R00209	C00022 + C00010 + C00003 <=> C00024 + C00011 + C00004 + C00080	-34.1	6.4
PWY-7003_09.2_R00209	C00022 + C00010 + C00003 <=> C00024 + C00011 + C00004 + C00080	-34.1	6.4
PWY-7003_10_R00238	2 C00024 <=> C00010 + C00332	26.2	1.7
PWY-7003_11_R01975	C00332 + C00004 + C00080 <=> C01144 + C00003	-60.4	2.4
PWY-7003_12_R03026	C01144 <=> C00877 + C00001	3.8	3.3
PWY-7003_13_R01175	C00877 + C01352 <=> C00136 + C00016	3.7	7.8
PWY-7003_14_R01173	C00136 + C00005 + C00080 <=> C01412 + C00010 + C00006	-38.1	5.6
PWY-7003_15_R03545	C01412 + C00005 + C00080 <=> C06142 + C00006	-70.0	4.8
<b>Total</b>		<b>-74.4 kJ/mol</b>	

Note-se ainda que nas décima primeira, décima terceira e décima quinta reacções é notório o comportamento anteriormente observado em *ReactionFingerprinter* e que leva à atribuição de estimativas baseadas em semelhanças entre reações com sentidos opostos. Quando acumulados, estes efeitos certamente apresentam uma forte contribuição para a grande diferença entre o saldo energético total obtido para CC e aquele que a Tabela 3.8. indica para *ReactionFingerprinter*.

O gráfico da Fig. 3.17. mostra a dispersão dos valores estimados por CC e por *ReactionFingerprinter*. Ao contrário do que era visível nos gráficos das Fig. 3.15. e Fig. 3.16., aqui não parece haver uma tendência tão marcada para que *ReactionFingerprinter* origine estimativas mais positivas e num intervalo de valores mais estreito do que CC.

Convém ainda notar uma vez mais que para esta via metabólica as primeiras nove reacções são consideradas em duplicado.



**Fig. 3.17.** – Dispersão das estimativas geradas por CC (a verde) e por *ReactionFingerprinter* (a vermelho) para as reacções da via de degradação do glicerol e produção de butanol. Aqui a tendência para que *ReactionFingerprinter* produza estimativas mais positivas e situadas numa faixa mais estreita não parece tão clara. Note-se que nesta via metabólica as primeiras nove reacções são consideradas em duplicado.

## Resultados e discussão

**Tabela 3.8.** - Estimativas para a energia livre de Gibbs,  $\Delta G_r^0$ , obtidas por *ReactionFingerprinter* para as reações que constituem a via metabólica da degradação do glicerol e produção de butanol. A cada estimativa está também associado o desvio-padrão, *StdDev*, da medição e valor do maior índice de semelhança com o qual foi obtida.

ReactionFingerprinter ID	KEGG Reaction	Melhor target	$\Delta_r G^0$ (kJ/mol)	StdDev	Índice de semelhança
PWY-7003_01.1_R01034	C00116 + C00003 <=> C00184 + C00004 + C00080	C00003 + C00116 <=> C00004 + C00184	66.40	1,27	1,000
PWY-7003_01.2_R01034	C00116 + C00003 <=> C00184 + C00004 + C00080	C00003 + C00116 <=> C00004 + C00184	66.40	1,27	1,000
PWY-7003_02.1_R01011	C00002 + C00184 <=> C00008 + C00111	C00002 + C00008 <=> C00020 + C03483	52.06	0	0,977
PWY-7003_02.2_R01011	C00002 + C00184 <=> C00008 + C00111	C00002 + C00008 <=> C00020 + C03483	52.06	0	0,977
PWY-7003_03.1_R01015	C00111 <=> C00118	C00001 + C00197 <=> C00009 + C00258	-17.69	0,29	0,797
PWY-7003_03.2_R01015	C00111 <=> C00118	C00001 + C00197 <=> C00009 + C00258	-17.69	0,29	0,797
PWY-7003_04.1_R01061	C00118 + C00009 + C00003 <=> C00236 + C00004 + C00080	C00003 + C00009 + C00118 <=> C00004 + C00236	53.27	3,54	1,000
PWY-7003_04.2_R01061	C00118 + C00009 + C00003 <=> C00236 + C00004 + C00080	C00003 + C00009 + C00118 <=> C00004 + C00236	53.27	3,54	1,000
PWY-7003_05.1_R01512	C00008 + C00236 <=> C00002 + C00197	C00002 + C00033 <=> C00008 + C00227	10.29	1,57	0,977
PWY-7003_05.2_R01512	C00008 + C00236 <=> C00002 + C00197	C00002 + C00033 <=> C00008 + C00227	10.29	1,57	0,977
PWY-7003_06.1_R01518	C00197 <=> C00631	C00001 + C00197 <=> C00009 + C00258	-17.69	0,29	0,852
PWY-7003_06.2_R01518	C00197 <=> C00631	C00001 + C00197 <=> C00009 + C00258	-17.69	0,29	0,852
PWY-7003_07.1_R00658	C00631 <=> C00074 + C00001	C00631 <=> C00001 + C00074	-3.31	0,67	1,000
PWY-7003_07.2_R00658	C00631 <=> C00074 + C00001	C00631 <=> C00001 + C00074	-3.31	0,67	1,000
PWY-7003_08.1_R00200	C00008 + C00074 <=> C00002 + C00022	C00002 + C00008 <=> C00020 + C03483	52.06	0	0,947
PWY-7003_08.2_R00200	C00008 + C00074 <=> C00002 + C00022	C00002 + C00008 <=> C00020 + C03483	52.06	0	0,947
PWY-7003_09.1_R00209	C00022 + C00010 + C00003 <=> C00024 + C00011 + C00004 + C00080	C00003 + C00010 + C00084 <=> C00004 + C00024	25.62	0,33	0,997
PWY-7003_09.2_R00209	C00022 + C00010 + C00003 <=> C00024 + C00011 + C00004 + C00080	C00003 + C00010 + C00084 <=> C00004 + C00024	25.62	0,33	0,997
PWY-7003_10_R00238	2 C00024 <=> C00010 + C00332	2 C00024 <=> C00010 + C00332	25.68	2,29	1,000
PWY-7003_11_R01975	C00332 + C00004 + C00080 <=> C01144 + C00003	C00003 + C05268 <=> C00004 + C05269	67.04	3,39	0,989
PWY-7003_12_R03026	C01144 <=> C00877 + C00001	C01144 <=> C00001 + C00877	3.80	0,90	1,000
PWY-7003_13_R01175	C00877 + C01352 <=> C00136 + C00016	C00016 <=> C01352	-77.55	0	0,924
PWY-7003_14_R01173	C00136 + C00005 + C00080 <=> C01412 + C00010 + C00006	C00003 + C01144 <=> C00004 + C00332	60.39	1,36	0,975
PWY-7003_15_R03545	C01412 + C00005 + C00080 <=> C06142 + C00006	C00006 + C00583 <=> C00005 + C00424	78.28	0,21	0,997
<b>Total</b>			<b>599.65 kJ/mol</b>		

## 3.4.3. BIOSÍNTESE DA VIOLACEÍNA

O processo de gerar estimativas para a energia livre de Gibbs para as reações da biossíntese de violaceína apresentou dificuldades por ambos os métodos. Em primeiro lugar, o processamento da primeira reação desta via por *CC* origina uma estimativa com um intervalo de confiança demasiado grande, colocando em causa a validade do valor obtido. Depois, a terceira reação origina um valor nulo, tanto para a estimativa da energia livre como para o respectivo intervalo de confiança. Não se tratando de uma reacção de isomerização, tudo leva a crer que esta reação pode ser incluída num dos grupos abordados nas secções 3.1.2. ou 3.1.3.. A **Tabela 3.9.** reflecte todos estes artefactos. No final, é essencialmente a contribuição das três últimas reações que leva a que o saldo energético estimado para a biossíntese da violaceína seja de  $-1945,5 \text{ kJ/mol}$ , sugerindo que esta é uma via altamente exergónica.

**Tabela 3.9.** - Estimativas para a energia livre de Gibbs,  $\Delta G^{\circ}$ , das reações da via de biossíntese de violaceína por *CC*, bem como para a energia total envolvida na totalidade da via. *Component Contribution* calcula ainda o erro padrão,  $Err^{Std}$ , associado a cada estimativa.

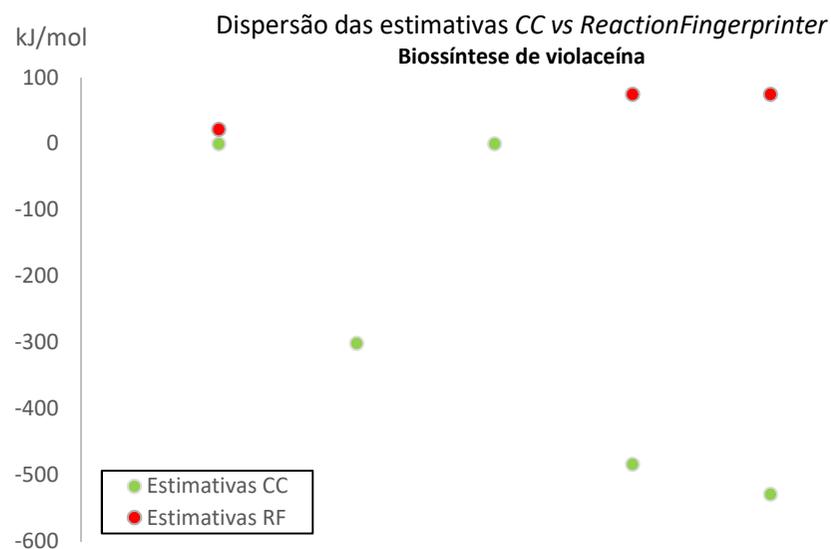
ReactionFingerprinter ID	KEGG Reaction	$\Delta_r G^{\circ}$ (kJ/mol)	$Err^{Std}$ (+/- kJ/mol)
PWY-7040_01	C00078 + C00007 <=> C91006 + C00027 + C00080	-124.2	98000.0
PWY-7040_02	2 C91006 + C00027 <=> C91007 + 2 C00001	-301.3	7.2
PWY-7040_03	C91007 + 2 C00080 <=> C21131 + C01342 + C00011	0.0	0.0
PWY-7040_04_R11158	C21131 + C00004 + C00007 + C00080 <=> C21134 + C00003 + C00001	-483.7	7.0
PWY-7040_05_R11373	C21134 + C00004 + C00007 + C00080 <=> C21135 + C00003 + C00001	-529.2	9.9
PWY-7040_06_R11136	C21135 + C00007 + C00080 <=> C21136 + C00011 + C00001	-507.1	13.9
<b>Total</b>		$-1945.5 \text{ kJ/mol}$	

A **Tabela 3.10.** indica os valores das estimativas obtidas por *ReactionFingerprinter* para três das cinco reações da biossíntese da violaceína. Novamente, *ReactionFingerprinter* não foi capaz de encontrar emparelhamentos *query-target* para três das reações com nenhum dos valores de limiar de dissimilaridade utilizados. Além disso, os valores obtidos para as reações que foram processadas divergem muito dos que foram obtidos por *CC*.

## Resultados e discussão

**Tabela 3.10.** - Estimativas para a energia livre de Gibbs,  $\Delta_r G^{\theta}$ , obtidas por *ReactionFingerprinter* para as reações que constituem a via metabólica da biossíntese da violaceína. A cada estimativa está também associado o desvio-padrão, *StdDev*, da medição e o valor do maior índice de semelhança com o qual foi obtida.

RF ID	KEGG Reaction	Melhor target	$\Delta_r G^{\theta}$ (kJ/mol)	StdDev	Índice de semelhança
PWY-7040_01	C00078 + C00007 <=> C91006 + C00027 + C00080	C00001 + C00078 <=> C00014 + C00022 + C00463	22.22	1,07	0,855
PWY-7040_02	2 C91006 + C00027 <=> C91007 + 2 C00001				
PWY-7040_03	C91007 + 2 C00080 <=> C21131 + C01342 + C00011				
PWY-7040_04_R11158	C21131 + C00004 + C00007 + C00080 <=> C21134 + C00003 + C00001	C00003 + C04367 <=> C00004 + C01244	75.05	0	0,888
PWY-7040_05_R11373	C21134 + C00004 + C00007 + C00080 <=> C21135 + C00003 + C00001	C00003 + C04367 <=> C00004 + C01244	75.05	0	0,877
PWY-7040_06_R11136	C21135 + C00007 + C00080 <=> C21136 + C00011 + C00001				



**Fig. 3.18.** – Dispersão das estimativas geradas por *CC* e por *ReactionFingerprinter* para as reações da via de biossíntese de violaceína. O reduzido número de reações para as quais *ReactionFingerprinter* gerou estimativas inviabiliza uma leitura conclusiva acerca da sua dispersão.

Uma vez mais, parece notar-se uma predisposição de *ReactionFingerprinter* para produzir estimativas dentro de um intervalo mais estreito e com tendência para valores mais positivos, como sugerido pelo gráfico da **Fig. 3.18**. Contudo, o reduzido número de emparelhamentos conseguidos por *ReactionFingerprinter* para as reacções desta via inviabiliza uma análise mais conclusiva às diferenças encontradas entre os resultados de ambos os métodos.



## **4. CONCLUSÕES E TRABALHO FUTURO**

---

---



Nesta dissertação é desenvolvido de raiz um método, *ReactionFingerprinter*, para estimar a variação da energia livre Gibbs em reacções bioquímicas. Este método é baseado na ideia de que os dados experimentais disponíveis na literatura podem ser aplicados segundo uma lógica de semelhanças estruturais e transformacionais para inferir dados acerca de reacções para as quais os parâmetros termodinâmicos são ainda desconhecidos, sendo os resultados obtidos comparados com os do método *Component Contribution*. Em termos gerais, *CC* aparenta proporcionar uma maior cobertura do que *ReactionFingerprinter* quando se considera a aplicação isolada de cada um dos métodos sobre a amostra de 200 reacções consideradas. Contudo, *ReactionFingerprinter* tem a capacidade de resolver alguns dos casos que não podem ser processados por *CC*, transimtindo a ideia de que *ReactionFingerprinter* poderá de facto trazer um acréscimo palpável ao número de reacções para as quais *CC* consegue fornecer estimativas. Em termos gerais, ambos os métodos poderão beneficiar da introdução de melhorias num futuro próximo, bem como da extensão das bases de dados de parâmetros termodinâmicos em reacções bioquímicas.

#### **4.1. APERFEIÇOAMENTOS EM COMPONENT CONTRIBUTION**

---

Em geral, *CC* é um método extremamente simples de utilizar e capaz de produzir resultados para um grande número de reacções numa questão de, no máximo, poucos minutos. É um método capaz de, em poucos instantes, carregar informações relativas a centenas de reacções de TEQR e conjugá-las com as tabelas de decomposição em grupos, usando-as em seguida para calcular estimativas. É, por isso, uma ferramenta que rapidamente se tornou uma referência mundial no que diz respeito à obtenção de estimativas para a energia livre de Gibbs.

Não obstante, as limitações expostas no **Capítulo 3** revelam ainda a existência de uma grande margem de melhoria, especialmente quanto à incapacidade de processar reacções com estruturas indefinidas, ou cuja decomposição não é possível no actual estado da arte da atribuição de energias livres de formação aos grupos definidos nas tabelas em que *CC* se baseia. O primeiro destes aspetos poderia ser resolvido recorrendo

## ***Conclusões e trabalho futuro***

a princípios conceptualmente simples, como são a substituição de grupos *-R* por elementos simples (átomos de hidrogénio ou pequenos grupos sem interferência significativa no cálculo de estimativas) ou a equiparação dos índices *n* das estruturas poliméricas a elementos neutros (os índices *n* são devidamente explicitados na estrutura dos ficheiros MDL e RXN, pelo que quando aparecem associados a compostos de ambos os lados da equação poderiam ser sinalizados e ignorados no cálculo final). Estas ideias em particular aparentam ser de implementação simples e permitiriam aumentar consideravelmente a capacidade de cobertura de *CC*. O caso das estruturas não decomponíveis é mais complexo e provavelmente beneficiaria mais da conjugação com um método como *ReactionFingerprinter*, que se centrasse essencialmente na transformação ocorrida, independentemente da complexidade das estruturas vizinhas. É o caso da reação ilustrada na Fig. 3.10.: apesar de incluir uma estrutura complexa e difícil de decompor, deveria ser possível identificar a sua presença de ambos os lados da equação. No final, o cancelamento mútuo dos termos respeitantes à estrutura não decomponível, deveria possibilitar um cálculo centrado naquelas que efetivamente são modificadas entre reagentes e produtos.

## **4.2. APERFEIÇOAMENTOS EM REACTION FINGERPRINTER**

---

O desenvolvimento de *ReactionFingerprinter* ao longo desta dissertação adveio do interesse científico em aplicar um mecanismo de estimativas baseado nos mesmos princípios que levaram ao desenvolvimento de *IGERS*<sup>[17]</sup>, capaz de proporcionar uma via de comparação de *CC* com uma metodologia assente na procura de similaridades químicas. Na sua formulação atual, *ReactionFingerprinter* apresenta uma cobertura inferior à que é oferecida por *CC*, mas com igual ou maior margem para introdução de melhorias.

Desde logo, é importante sublinhar que este método depende atualmente da ferramenta comercial *JChem*®, cuja documentação não especifica a ponderação que é atribuída aos vários critérios utilizados no cálculo de índices de semelhança. Convém também salientar que estes critérios vão além daqueles que são utilizados por *IGERS*: enquanto este último assenta principalmente na definição de atributos químicos e na

sua existência – ou ausência – em cada um dos lados da reação, a *Similarity Search* de *JChem*® procura ainda semelhanças entre os átomos que se encontram na vizinhança da transformação e entre as estruturas químicas das próprias moléculas que tomam parte na reação. Esta maior especificidade leva a crer que os índices de semelhança calculados poderão também ser mais precisos do que os calculados por *IGERS*. Independentemente disso, seria importante em abordagens futuras saber ao certo como é que estes critérios são integrados no cálculo do índice final, para que se perceba se é colocada maior ênfase na questão estrutural ou transformacional. Entre outros aspetos, esse conhecimento poderia ajudar a afinar o conceito daquilo que é um limiar de dissemelhança aceitável quando se pretende gerar estimativas para a energia livre de Gibbs. A reação ilustrada na **Fig. 3.9.** poderá ser um exemplo: o tipo de reação retratado está longe de ser invulgar, mas é possível que a peculiaridade da estrutura molecular envolvida absorva a vulgaridade da transformação propriamente dita; o cálculo de um índice mais centrado nas alterações dos átomos envolvidos na reação poderia significar um aumento no número de emparelhamentos eficazes.

Relativamente a *ReactionFingerprinter*, é ainda fundamental corrigir a incapacidade de detetar a existência de eventuais desequilíbrios químicos nas reações analisadas, bem como a atribuição de energias com o sinal simétrico sempre que o emparelhamento é feito com *targets* que estão descritos no sentido inverso ao da transformação representada pela *query*. Especialmente o segundo caso permitiria a correção e aproximação à realidade de muitas das estimativas calculadas.

#### **4.3. EXTENSÃO DAS BASES DE DADOS SOBRE A TERMODINÂMICA DE REAÇÕES BIOQUÍMICAS**

---

Independentemente dos méritos e limitações das metodologias abordadas, há um outro ponto fundamental a considerar para que se possam introduzir melhorias nas estimativas de qualquer parâmetro termodinâmico: o estudo de mais reações bioquímicas e a inclusão da informação obtida em bases de dados como *TECR*. É certo que este trabalho tem início com o pressuposto de que são muito poucas as reações

## ***Conclusões e trabalho futuro***

metabólicas para as quais existem estudos termodinâmicos, e que os mesmos envolvem processos fastidiosos e pouco adequados a gerar, em tempo útil, informações para um grande número de reações. Todavia, a análise levada a cabo para os dois métodos estudados leva a crer que o acréscimo de um pequeno número de reações às bases de dados poderia significar um grande aumento na capacidade de cobertura dos métodos, desde que esse pequeno número esteja concentrado nos tipos de reações certas.

Ambos os métodos revelam limitações em casos para os quais os dados experimentais escasseiam: em *CC* faltam dados para aplicar *RC* ou *CC* (ou ambos) a mais reações, enquanto que em *ReactionFingerprinter* faltam *targets* mais variados para a obtenção de um maior número de emparelhamentos para os limiares de dissemelhança definidos. Por exemplo, sabendo que muitos destes casos resultam de reações como as exemplificadas na secção 3.1.3., é legítimo supor que estudos centrados na termodinâmica de reações com estruturas moleculares menos vulgares poderiam acarretar um acréscimo significativo no conhecimento disponível, tanto em termos das energias de formação de novos compostos como em termos da definição de possíveis decomposições das suas estruturas. Este conhecimento adicional poderia ainda ter impacto ao nível da variedade dos *targets* disponíveis para comparações por *ReactionFingerprinter*, ou mesmo eliminar algumas das elevadas incertezas associadas a reacções analisadas por *CC*.

A extensão das bases de dados poderia nem depender necessariamente da produção de um grande número de estudos experimentais. É recorrente a publicação de artigos que referem de forma detalhada o estudo termodinâmico de sistemas biológicos menos comuns, havendo a possibilidade da integração dos seus resultados nas tabelas já existentes.<sup>[59, 60, 61, 62]</sup> As bases de dados poderiam ainda ser aperfeiçoadas com dados provenientes de outro tipo de métodos, como são os baseados em Mecânica Quântica e abordagens híbridas *QM-MM/MD*. Apesar de estes carecerem do carácter experimental, as suas estimativas são frequentemente suficientemente próximas da realidade para poderem ser incluídos nas tabelas existentes. Além disso, são métodos que poderiam ser utilizados no sentido de validar os dados disponíveis para reacções cujos parâmetros conhecidos resultam de medições únicas ou realizadas sob condições pouco claras.

#### 4.4. TRABALHO FUTURO

---

Atendendo aos resultados produzidos neste trabalho, a curto prazo as principais melhorias a serem introduzidas nas metodologias descritas centrar-se-iam essencialmente em *ReactionFingerprinter*. Desde logo porque este é um método moroso, que implica a chamada de um processo de comparação para cada *query* e para cada *target* individualmente. A reorganização do algoritmo de maneira a aglomerar, por exemplo, a comparação de uma *query* com todos os *targets* numa única chamada poderia diminuir consideravelmente o tempo necessário para computação. Consequentemente, poder-se-ia dispor em tempo útil de estimativas para um maior número de reações. (Na atual formulação, e num computador de secretária comum, são necessárias cerca de 50 horas para que *ReactionFingerprinter* calcule estimativas para a amostra de 200 reações estudadas.) A preparação do código para execução num *cluster* de processadores significaria uma diminuição ainda mais substancial no tempo de computação, uma vez que permitiria alocar recursos em paralelo e assim correr um grande número de comparações em simultâneo. A redução no tempo de computação poderia significar, por exemplo, o cálculo de estimativas para as cerca de 10 000 reações que constam de KEGG, oferecendo uma amostra mais vasta e mais próxima daquilo que é o número e variedade de reações metabólicas descritas na literatura. Outra correção perfeitamente exequível a curto prazo seria a introdução de um filtro para reações que fujam ao princípio de conservação da massa e a atribuição dos valores de  $\Delta_rG$  simétricos aquando do emparelhamento entre reações com sentidos opostos.

A médio prazo seria ainda possível introduzir um maior controlo na ponderação atribuída aos critérios utilizados na definição dos índices de semelhança, permitindo aproximar o índice resultante a uma lógica mais estrutural ou mais transformacional, consoante a natureza e objetivos da análise efectuada. Também *CC* poderia ser adaptado para assinalar convenientemente os casos de isomerização, fazendo a distinção automática entre estes e aqueles aos quais é atribuída artificialmente uma variação nula na energia livre por incapacidade de processamento. Quando se pretende analisar um elevado número de reações – na ordem dos milhares – esta sinalização proporcionaria uma capacidade acrescida para validar resultados de uma forma automatizada. O processamento de grupos *-R* e estruturas poliméricas de acordo com o sugerido na

### ***Conclusões e trabalho futuro***

secção 4.1. é outro exemplo interessante, e permitiria um incremento considerável na capacidade de processamento de *CC*.

A introdução das melhorias sugeridas permitiria aprofundar melhor a questão da complementaridade entre os dois métodos. Em última instância, ficaria traçado de forma mais concreta o molde sobre o qual poderia assentar uma ferramenta mais completa para o estudo termodinâmico das reações bioquímicas.

## REFERÊNCIAS

---

- [1] Cameron, D.E., Bashor, C.J., Collins, J.J. (2014). *A brief history of synthetic biology*, Nature Reviews – Microbiology **12**, 381-390.
- [2] Vargas-Maya Naurú, I., Bernardo, F., *Escherichia coli – Recent advances on physiology, pathogenesis and biotechnological applications, Chapter 13 – Escherichia coli as a model organism and its application in biotechnology*, INTECH, July 12, 2017.
- [3] Karatia, H., Vilaprinyo, E., Sorribas, A., Alves, R. (2011). *Saccharomyces cerevisiae as a model organism: a comparative study*, PLoS ONE **6**(2).
- [4] Lee, J.W., Kim, T.Y., Jang, Y.S., Choi, S., Lee, S.Y. (2011). *Systems metabolic engineering for chemicals and materials*, Trends in Biotechnology **29**(8), 370-378.
- [5] Liu, L., Redden, H., Alper, H.S. (2013). *Frontiers of yeast metabolic engineering: diversifying beyond ethanol and Saccharomyces*, Current Opinion in Biotechnology **24**, 1023-1030.
- [6] Yang, L., Cluett, W.R., Mahadevan, R. (2011). *EMILiO: A fast algorithm for genome-scale strain-design*, Metabolic Engineering **13**, 272-281.
- [7] Machado, D., Herrgard, M.J. (2015). *Co-evolution of strain design methods based on flux balance and elementary mode analysis*, Metabolic Engineering Communications **2**, 85-92.
- [8] Long, M.R., Ong, W.K., Reed, J.L. (2015). *Computational methods in metabolic engineering for strain design*, Current Opinion in Biotechnology **34**, 135-141.
- [9] King, Z.A., Lloyd, C.J., Feist, A.M., Palsson, B.O. (2015). *Next-generation genome-scale models for metabolic engineering*, Current Opinion in Biotechnology **35**, 23-29.
- [10] Almeida, P.F., Moreira, R.S., Almeida, R.C.C., Guimarães, A.K., Carvalho, A.S., Quintella, C., Esperidiã, M.C.A., Taft, C.A. (2004). *Selection and application of microorganisms to improve oil recovery*, Engineering in Life Sciences **4**(4), 319-325.
- [11] Escobar-Niño, A., Luna, C., Luna, D., Marcos, A.T., Cánovas, D., Mellado, E. (2014). *Selection and characterization of biofuel-producing bacteria isolated from vegetable oil-rich wastes*, PLoS ONE **9**(8), 1-12.
- [12] Khetkorn, W., Khanna, N., Incharoensakdi, A., Lindblad, P. (2013). *Metabolic and genetic engineering of cyanobacteria for enhanced hydrogen production*, Biofuels **4**(5), 535-561.
- [13] Carpine, R., Du, W., Olivieri, G., Pollio, A., Hellingwerf, K.J., Marzocchella, A., dos Santos, F.B. (2017). *Genetic engineering of Synechocystis sp. PCC6803 for poly-β-hydroxybutyrate overproduction*, Algal Research **25**, 117-127.
- [14] Fisher, A.K., Freedman, B.G., Bevan, D.R., Senger, R.S. (2014). *A review of metabolic and enzymatic engineering strategies for designing and optimizing performance of microbial cell*

*factories*, Computational and Structural Biotechnology Journal **11**, 91-99.

[15] Alberty, R.A. (2004). *A short history of the thermodynamics of enzyme-catalyzed reactions*, The Journal of Biological Chemistry **279**(27), 27831-27836.

[16] Fleming, R.M.T., Thiele, I., Nasheuer, H.P. (2009). *Quantitative assignment of reaction directionality in constraint-based models of metabolism: application to Escherichia coli*, Biophysical Chemistry **145**, 47-56.

[17] Rother, K., Hoffman, S., Bulik, S., Hoppe, A., Gasteiger, J., Holzhütter, H.G. (2010). *IGERS: Inferring Gibbs energy changes from reaction similarities*, Biophysical Journal **98**, 2478-2486.

[18] Noor, E., Bar-Even, A., Flamholz, A., Lubling, Y., Davidi, D., Milo, R. (2012). *An integrated open framework for thermodynamics of reactions that combines accuracy and coverage*, Bioinformatics **28**(15), 2037-2044.

[19] Haraldsdóttir, H.S., Thiele, I., Fleming, R.M.T. (2012). *Quantitative assignment of reaction directionality in a multicompartmental human metabolic reconstruction*, Biophysical Journal **102**, 1703-1711.

[20] Noor, E., Haraldsdóttir, H.S., Milo, R., Fleming, R.M.T. (2013). *Consistent estimation of Gibbs energy using Component Contributions*, PLOS Computational Biology **9**(7).

[21] Voet, D.J., Voet, J.G., Pratt, C.W., *Principles of Biochemistry, Chapter 1 – Life, Cells. and Thermodynamics*, John Wiley & Sons, Inc., 2018.

[22] Mavrovouniotis, M.L. (1991). *Estimation of standard Gibbs energy changes of biotransformations*, The Journal of Biological Chemistry **266**(22).

[23] Glaser, L., Brown, D.H. (1955). *Purification and properties of D-Glucose-6-phosphate dehydrogenase*, The Journal of Biological Chemistry **216**, 67-79.

[24] Casazza, J.P., Veech, R.L. (1986). *The interdependence of glycolytic and pentose cycle intermediates in ad libitum fed rats*, The Journal of Biological Chemistry **261**(2), 690-698.

[25] Goldberg, R.N., Tewari, Y.B., Bell, D., Fazio, K., Anderson, E. (1993). *Thermodynamics of enzyme catalyzed reactions: Part 1. Oxidoreductases*, Journal of Physical and Chemical Reference Data **22**(2), 515-582.

[26] Goldberg, R.N., Tewari, Y.B. (1994). *Thermodynamics of enzyme-catalyzed reactions: part 2. Transferases*, Journal of Physical and Chemical Reference Data **23**(4), 547-617.

[27] Goldberg, R.N., Tewari, Y.B. (1994). *Thermodynamics of enzyme-catalyzed reactions: Part 3. Hydrolases*, Journal of Physical and Chemical Reference Data **23**(6), 1035-1103.

[28] Goldberg, R.N., Tewari, Y.B. (1995). *Thermodynamics of enzyme-catalyzed reactions: Part 4. Lyases*, Journal of Physical and Chemical Reference Data **24**(5), 1669-1698.

[29] Goldberg, R.N., Tewari, Y.B. (1995). *Thermodynamics of enzyme-catalyzed reactions: Part 5. Isomerases and Ligases*, Journal of Physical and Chemical Reference Data **24**(6), 1765-1801.

[30] Benson, S.W., Buss, J.H. (1958). *Additivity rules for the estimation of molecular properties*.

*Thermodynamic properties*, The Journal of Chemical Physics **29**(3), 546-572.

[31] Domalski, E.S., Hearing, E.D. (1988). *Estimation of the thermodynamic properties of hydrocarbons at 298.15 K*, Journal of Physical and Chemical Reference Data **17**(4), 1637-1678.

[32] Mavrovouniotis, M.L., Bayol, P., Lam, T.K.M., Stephanopoulos, G., Stephanopoulos, G. (1988). *A group contribution method for the estimation of equilibrium constants for biochemical reactions*, Biotechnology Techniques **2**(1), 23-28.

[33] Mavrovouniotis, M.L. (1991). *Estimation of standard Gibbs energy changes of biotransformations*, The Journal of Biological Chemistry **266**(22), 14440-14445.

[34] Jankowski, M.D., Henry, C.S., Broadbelt, L.J., Hatzimanikatis, V. (2008). *Group contribution method for thermodynamic analysis of complex metabolic networks*, Biophysical Journal **95**(3), 1487-1499.

[35] Feist, A.M., Henry, C.S., Reed, J.L., Krummenacker, M., Joyce, A.R., Karp, P.D., Broadbelt, L.J., Hatzimanikatis, V., Palsson, B.O. (2007). *A genome-scale metabolic reconstruction for Escherichia coli k-12 MG1655 that accounts for 1260 ORFs and thermodynamic information*, Molecular Systems Biology **3**(121).

[36] Krebs, H.A, Kornberg, H.L., Burton, K. (1957). *A survey of the energy transformations in living matter*, Ergebnisse der physiologie biologischen chemie und experimentellen pharmakologie, **49**(1), 212-298.

[37] Thauer, R.K., Jungermann, K., Decker, K. (1977). *Energy conservation in chemotrophic anaerobic bacteria*, Bacteriological Reviews **41**(1), 100-180.

[38] Alberty, R.A. (2003). *Thermodynamics of biochemical reactions*, John Wiley & Sons, Hoboken N.J..

[39] Alberty, R.A. (2006). *Biochemical thermodynamics: applications of Mathematica (Methods of Biochemical Analysis)*, John Wiley & Sons, Hoboken N.J..

[40] Jinich, A., Rappoport, D., Dunn, I., Sanchez-Lengeling, B., Olivares-Amaya, R., Noor, E., Bar-Even, A., Aspuru-Guzik, A. (2014). *Quantum-chemical approach to estimating the thermodynamics of metabolic reactions*, Scientific Reports **4**(7022), 1-6.

[41] Aneesh, M.H. (2012). *A theoretical study on the regioselectivity of electrophilic reactions of heterosubstituted allyl systems*, Chapter I – Ab initio methods in computational chemistry, Department of Chemistry, University of Calicut.

[42] Becker, O.M., MacKerell Jr, A.D., Roux, B., Watanabe, M. (2001). *Computational Biochemistry and Biophysics*, Chapter 3 – Dynamics Methods, Marcel Dekker, Inc., New York N.Y..

[43] Tuñón, I., Silla, E., Millot, C., Martins-Costa, M.T.C., Ruiz-López, M.F. (1998). *Intramolecular proton transfer of glycine in aqueous solution using Quantum mechanics-Molecular mechanics simulations*, The Journal of Physical Chemistry A **102**(45), 8673-8678.

[44] Leach, A.R. (2001). *Molecular Modelling – Principles and Applications*, 2nd Edition, Pearson

Education Limited, Edinburgh Gate, Harlow, England.

[45] Becker, O.M., MacKerell Jr, A.D., Roux, B., Watanabe, M. (2001). *Computational Biochemistry and Biophysics*, Chapter 11 – Computer simulation of biochemical reactions with QM-MM methods, Marcel Dekker, Inc., New York N.Y..

[46] Friesner, R.A. (2005). *Ab initio quantum chemistry: methodology and applications*. PNAS **102**(19), 6648-6653.

[47] Monard, G., Merz, K.M. Jr. (1999). *Combined Quantum Mechanical/Molecular Mechanical methodologies applied to biomolecular systems*, Accounts of Chemical Research **30**(10), 904-911.

[48] Rod, T.H., Ryde, U. (2005). *Quantum mechanical free energy barrier for an enzymatic reaction*. Physical Review Letters **94**(13).

[49] Hu, H., Yang, W. (2008). *Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanics/molecular mechanics methods*. Annual Review of Physical Chemistry **59**, 573-601.

[50] van der Kamp, M.W., Mulholland, A.J. (2013). *Combined Quantum Mechanics/Molecular Mechanics (QM/MM) methods in computational enzymology*. Biochemistry **52**, 2708-2728.

[51] Goldberg, R.N., Tewari, Y.B., Bhat, T.N. (2004). *Thermodynamics of enzyme-catalyzed reactions – a database for quantitative biochemistry*, Bioinformatics Applications Note **20**(16), 2874-2877.

[52] Kanehisa, M., Goto, S. (2000). *KEGG: Kyoto Encyclopedia of Genes and Genomes*, Nucleic Acids Research **28**(1), 27-30.

[53] ChemAxon Documentation – MDL MOLfiles, RGfiles, SDfiles, Rxnfiles, RDfiles formats (Last modified on jan 04, 2018), <https://docs.chemaxon.com/display/docs/MDL+MOLfiles%2C+RGfiles%2C+SDfiles%2C+Rxnfiles%2C+RDfiles+formats>, Outubro 2018

[54] ChemAxon Documentation – Similarity search (Last modified on ago 03, 2018), <https://docs.chemaxon.com/display/docs/Similarity+search>, Outubro 2018

[55] ChemAxon Documentation – Reaction fingerprint RF (Last modified on fev 05, 2018), <https://docs.chemaxon.com/display/docs/Reaction+fingerprint+RF>, Outubro 2018

[56] ChemAxon Documentation – General Search Options (Last modified on set 19, 2016), <https://docs.chemaxon.com/display/docs/General+search+options#Generalsearchoptions-searchtype>, Outubro 2018

[57] ChemAxon Documentation – Similarity Specific Search Options (Last modified on set 19, 2016), <https://docs.chemaxon.com/display/docs/Similarity+specific+search+options#Similarityspecificsearchoptions-similaritythreshold>, Outubro 2018

- [58] ChemAxon Documentation – Hitdisplay specific search options (Last modified on set 19, 2016),  
<https://docs.chemaxon.com/display/docs/Hitdisplay+specific+search+options#Hitdisplayspecificsearchoptions-similarity>, Outubro 2018
- [59] Li, X., Dash, R.K., Pradhan, R.K., Qi, F., Thompson, M., Vinnakota, K.C., Wu, F., Yang, F., Beard, D.A. (2010). *A database of thermodynamic quantities for the reactions of glycolysis and the tricarboxylic acid cycle*, The Journal of Physical Chemistry B **114**(49), 16068-16082.
- [60] Li, X., Wu, F., Qi, F., Beard, D.A. (2011). *A database of thermodynamic properties of the reactions of glycolysis, the tricarboxylic acid cycle, and the pentose phosphate pathway*, Database: The Journal of Biological Databases and Curation **2011**.
- [61] Vinnakota, K., Kemp, M.L., Kushmerick, J. (2006). *Dynamics of muscle glycogenolysis modeled with pH time course computation and pH-dependent reaction equilibria and enzyme kinetics*, Biophysical Journal **91**(4), 1264-1287.
- [62] Amend, J.P., Shock, E.L. (2001). *Energetics of overall metabolic reactions of thermophilic and hyperthermophilic Archaea and Bacteria*, FEMS Microbiology Reviews **25**, 175-243.
- [63] Sprenger, G.A. (2007). *Aromatic Amino Acids*, Amino Acid Biosynthesis: Pathways, Regulation and Metabolic Engineering, Springer, 93-127.
- [64] Patel, N., Pierson, D.L., Jensen, R.A. (1977). *Dual enzymatic routes to L-tyrosine and L-phenylalanine via pretyrosine in Pseudomonas aeruginosa*, The Journal of Biological Chemistry **252**(16), 5839-5846.
- [65] Jung, E., Zamir, L.O., Jensen, R.A. (1986). *Chloroplasts of higher plants synthesize L-phenylalanine via L-arogenate*, Proceedings of the National Academy of Sciences **83**, 7231-7235.
- [66] Zhou, H., Liao, X., Wang, T., Du, G., Chen, J. (2010). *Enhanced L-phenylalanine biosynthesis by co-expression of pheA<sup>br</sup> and aroF<sup>wt</sup>*, Bioresource Technology **101**, 4151-4156.
- [67] Howles, P.A., Sewalt, V.J.H., Paiva, N.L., Elkind, Y., Bate, N.J., Lamb, C., Dixon, R.A. (1996). *Overexpression of L-phenylalanine ammonia-lyase in transgenic tobacco plants reveals control points for flux into phenylpropanoid biosynthesis*, Plant Physiology **112**, 1617-1624.
- [68] Achnine, L., Blancaflor, E.B., Rasmussen, S., Dixon, R.A. (2004). *Colocalization of L-phenylalanine ammonia-lyase and cinnamate 4-hydroxylase for metabolic channeling in phenylpropanoid biosynthesis*, The Plant Cell **16**, 3098-3109.
- [69] Lapadatescu, C., Giniès, C., Le Quéré, J.L., Bonnarme, P. (2000). *Novel scheme for biosynthesis of aryl metabolites from L-phenylalanine in the fungus Bjerkandera adusta*, Applied and Environmental Microbiology **66**(4), 1517-1522.
- [70] Wittstock, U., Halkier, B.A. (2000). *Cytochrome P450 CYP79A2 from Arabidopsis thaliana L. catalyzes the conversion of L-phenylalanine to phenylacetaldoxime in the biosynthesis of benzylglucosinolate*, The Journal of Biological Chemistry **275**(19), 14659-14666.

- [71] MetaCyc Pathway (PWY-6628) – Superpathway of L-Phenylalanine Biosynthesis, <https://metacyc.org/META/NEW-IMAGE?type=NIL&object=PWY-6628&redirect=T>, Outubro 2018
- [72] Antoni, D., Zverlov, V.V., Schwarz, W.H. (2007). *Biofuels from microbes*, Applied Microbiology and Biotechnology **77**(1), 23-35.
- [73] Ezeji, T.C., Qureshi, N., Blaschek, H.P. (2007). *Bioproduction of butanol from biomass: from genes to bioreactors*, Current Opinion in Biotechnology **18**, 220-227.
- [74] Dasari, M.A., Kiatsimkul, P.P., Sutterlin, W.R., Suppes, G.J. (2005). *Low-pressure hydrogenolysis of glycerol to propylene glycol*, Applied Catalysis A: General **281**, 225-2331.
- [75] Miyazawa, T., Kusunoki, Y., Kunimori, K., Tomishige, K. (2006). *Glycerol conversion in the aqueous solution under hydrogen over Ru/C + na ion-exchange resin and its reaction mechanism*, Journal of Catalysis **240**, 213-221.
- [76] Johnson, D.T., Taconi, K.A. (2007). *The glycerin glut: options for the value-added conversion of crude glycerol resulting from biodiesel production*, Environmental Progress **26**(4), 338-348.
- [77] MetaCyc Pathway (PWY-7003) – Glycerol degradation to butanol, <https://metacyc.org/META/NEW-IMAGE?type=NIL&object=PWY-7003&redirect=T>, Outubro 2018
- [78] Pemberton, J.M., Vincent, K.M., Penfold, R.J. (1991). *Cloning and heterologous expression of the violacein biosynthesis gene cluster from Chromobacterium violaceum*, Current Microbiology **22**, 355-358.
- [79] August, P.R., Grossman, T.H., Minor, C., Draper, M.P., MacNeil, I.A., Pemberton, J.M., Call, K.M., Holt, D., Osburne, M.S. (2000). *Sequence analysis and functional characterization of the violacein biosynthetic pathway from Chromobacterium violaceum*, Journal of Molecular Microbiology and Biotechnology **2**(4), 513-519.
- [80] Antônio, R.V., Creczynski-Pasa, T.B. (2004). *Genetic analysis of violacein biosynthesis by Chromobacterium violaceum*, Genetics and Molecular Research **3**(1), 85-91.
- [81] Sánchez, C., Braña, A.F., Méndez, C., Salas, J.A. (2006). *Reevaluation of the violacein biosynthetic pathway and its relationship to indolocarbazole biosynthesis*, ChemBioChem **7**, 1231-1240.
- [82] Hoshino, T. (2011). *Violacein and related tryptophan metabolites produced by Chromobacterium violaceum: biosynthetic mechanism and pathway for construction of violacein core*, Applied Microbiology and Biotechnology **91**, 1463-1475.
- [83] MetaCyc Pathway (PWY-7040) – Violacein biosynthesis, <https://metacyc.org/META/NEW-IMAGE?type=NIL&object=PWY-7040&redirect=T>, Outubro 2018