



Instituto de Química  
Programa de Pós-Graduação em Química  
Tese de Doutorado

## **O Uso da Inteligência Artificial e Outras Tecnologias na Caracterização de Fenômenos Químicos e Biológicos**

HUGO GONTIJO MACHADO

Tese de Doutorado apresentada ao Instituto de  
Química da Universidade de Brasília como  
requisito para obtenção do título de Doutor em  
Química

**ORIENTADOR: Prof. Dr. Kleber Carlos Mundim**

Brasília, DF

2022

HUGO GONTIJO MACHADO

**O Uso da Inteligência Artificial e Outras Tecnologias na  
Caracterização de Fenômenos Químicos e Biológicos**

Tese de Doutorado apresentada ao Instituto de  
Química da Universidade de Brasília como  
requisito para obtenção do título de Doutor em  
Química

**ORIENTADOR: Prof. Dr. Kleber Carlos Mundim**

Brasília, DF

2022

## AGRADECIMENTOS

Ao Prof. Dr. Kleber Mundim, o qual sempre admirei e pude ter a honra de desenvolver este projeto junto. Agradeço por ter confiado em mim na execução deste trabalho, sempre fomentando as mais instigantes discussões, além de me dar a oportunidade de ter contato com grandes pesquisadores. Foi um prazer todos estes anos de discussões e trabalho e espero que continuemos sempre trabalhando juntos.

Ao Prof. Dr. Valter Henrique Carvalho-Silva, com quem trabalhei desde a graduação e que me apresentou o universo da química teórica e computacional, além de sempre me receber e contribuir tanto no desenvolvimento deste trabalho quanto na minha formação humana.

A todos os professores e colaboradores que contribuíram para o desenvolvimento deste trabalho: Solange Baraldi, Fernando Vieira, Claudene Barros e Wender Alves.

Ao meu amigo e colega Flávio Sanches, que esteve ao meu lado desde a graduação e em todos os cursos acadêmicos e trabalhos científicos realizados até aqui. Agradeço também a todos os colegas dos grupos de pesquisa de Brasília: Daniel Scalabrini, Matheus, Sandro, Thiago e Prof. Dr. Heibbe Cristian; e do grupo de pesquisa de Anápolis: Priscila Gomes, Yago Francisco, Lilian Carvalho, Allane Carvalho; que de alguma forma contribuíram para realização deste trabalho. Agradeço também a todos os injustiçados, pessoas que me ajudaram neste período e que esqueci de citar aqui.

A pessoas muito queridas que passaram por minha vida: Rebeca Tayanne e Carppio de Moraes. Muito obrigado pelos anos de apoio e o suporte essencial para realização deste doutorado na cidade de Brasília.

À minha avó Dirce Maria de Jesus, pelos conselhos e a toda minha família por estarem sempre ao meu lado nesta jornada.

À minha companheira Eduarda Coelho, uma pessoa extraordinária que conheci durante esta jornada e que se manteve ao meu lado nos momentos mais difíceis e nos mais felizes, contribuindo para que eu pudesse estar aqui hoje.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro e bolsa concedida.

À agência CNPQ pelo suporte financeiro e a UEG e UnB pela utilização das unidades de processamento de alta performance.

Em especial aos meus pais, Ana Lúcia Gontijo e Rinaldo Vieira Machado, que sempre acreditaram em mim e me deram todo o apoio possível sem medir esforços para que eu tivesse uma formação acadêmica. Sem dúvida, eu jamais teria conseguido chegar até aqui sem este apoio. Espero algum dia estar próximo das grandes pessoas que eles representam para mim.

## RESUMO

Modelos de aprendizagem estatística que utilizam métodos de *Machine Learning* (ML), como o *Generalized Simulated Annealing* (GSA), vêm sendo utilizados há décadas na química para diversas tarefas. Munido destas técnicas, nossos grupos de pesquisa têm se dedicado ao estudo cinético e mecanístico de processos reativos e não reativos aliado ao desenvolvimento de novos modelos e técnicas. Com o aumento da capacidade de processamento dos computadores e com advento de métodos sofisticados de ML, como o *deep learning*, vem se tornando cada vez mais necessário a criação de novas técnicas de análise bem como o desenvolvimento de plataformas que sistematizem essas técnicas. Neste sentido, o presente trabalho realiza três propostas de modelos teóricos que serão apresentadas juntamente ao desenvolvimento de interfaces gráficas e intuitivas: *i*) propomos aqui a função da transitividade: uma nova perspectiva sob o gráfico de *Arrhenius* que é capaz de linearizar curvaturas não *Arrhenius* e observar mudanças de regime. Sua aplicação linearizou a curvatura do processo de relaxamento do carbonato de propileno revelando uma quebra do regime à temperatura de 198K. Além disso, o código *Transitivity* sistematiza ferramentas para lidar com a cinética de processos não *Arrhenius*, fornecendo opções para estimativa da constante cinética reacional em fases líquida e gasosa além de realizar ajuste de modelos cinéticos a partir de dados experimentais utilizando o GSA. Uma análise cinética da síntese de chalconas foi realizada usando o *Transitivity*. *ii*) propomos também o modelo generalizado do coeficiente da temperatura  $Q_{d10}$ , utilizado para avaliar o grau de dependência de processos biológicos com a temperatura. O modelo foi aplicado a uma série de casos da literatura recente: a qualidade dos ajustes reforçou a hipótese inicial de que o coeficiente da temperatura deve ser descrito através de exponenciais deformadas ( $Q_{d10}$ ) ao invés do modelo usual de *Arrhenius* ( $Q_{10}$ ). Além disso, o modelo padrão  $Q_{10}$  é um caso particular do modelo proposto, pois é recuperado quando o parâmetro deformado é igual a zero, validando assim sua justificativa geral. A interface  $Q_{d10} - GSA$  desenvolvida estima parâmetros a partir de dados experimentais, traçando um paralelo entre o modelo usual  $Q_{10}$  e o modelo proposto  $Q_{d10}$ . *iii*) propomos também modelos de *deep learning* baseados em redes neurais artificiais que aprendem com um banco de dados espectral para realizar inferências sobre propriedades estruturais de moléculas e realizar reconstruções dos espectros. As classificações realizadas pelos modelos ResNet34 e o modelo tabular apresentaram um bom desempenho utilizando espectros de infravermelho e massa, com valores de *F1-Score* entre 0.83 e 0.88 e acurácias de 96%. Os modelos de reconstrução apresentaram um erro quadrático médio de 0.005 e 0.007 na reconstrução dos espectros de massa e infravermelho, respectivamente. Com isso, o desenvolvimento destes modelos sugere que redes neurais são adequadas para a busca de novas técnicas de análise espectral e representam um passo no sentido da criação de técnicas de análise autônomas. Por fim, as interfaces intuitivas e amigáveis têm o intuito de disseminar os novos conceitos propostos e os autores esperam que possam ser utilizadas como poderosas ferramentas de pesquisa e didática no ensino de termodinâmica, cinética de processos e assuntos relacionados.

**PALAVRAS-CHAVE:** teoria deformada de *Arrhenius*, cinética química, coeficiente da temperatura  $Q_{10}$ , processos biológicos, *machine learning*, desenvolvimento de *software*.

## ABSTRACT

Statistical learning models that use machine learning (ML) methods, like the Generalized Simulated Annealing (GSA), have been used for decades in chemistry for a variety of tasks. With these techniques, our research groups are dedicated to the kinetic and mechanistic study of reactive and non-reactive processes allied to the development of new models and techniques. With the current increase in the processing capacity of computers and the advent of sophisticated methods of ML, such as deep learning, it has become increasingly necessary to create new analysis techniques as well as the development of platforms and tools that systematize these techniques. Accordingly, the present work makes three proposals for theoretical models that will be presented with the development of codes for their use, with graphical and intuitive interfaces: *i)* We propose here the transitivity function: a new perspective under Arrhenius plot that is able to linearize non-Arrhenius curvatures and thus observe regime changes. Its application was able to linearize the curvature of the propylene carbonate relaxation process and revealed a break in the regime at the temperature of 198K. In addition, the Transitivity code systematizes tools to deal with the kinetics of non-Arrhenius processes, providing options for phenomenological estimation of the reaction kinetic constant in liquid and gas phases and perform model fitting with experimental kinetic data using GSA. A kinetic analysis of the synthesis of chalcones was performed using Transitivity *ii)* we also propose the generalized model of the temperature coefficient  $Q_{d10}$ , used to evaluate the degree of dependence of biological processes as a function of temperature. The model was applied to a series of cases from the recent literature: the quality of the fits reinforced the initial hypothesis that the temperature coefficient must be described using deformed exponentials ( $Q_{d10}$ ) instead of the usual Arrhenius model ( $Q_{10}$ ). Furthermore, the standard model  $Q_{10}$  is a particular case of the proposed model, as it is retrieved when the deformed parameter is equal to zero, thus validating its general justification. The  $Q_{d10}$ -GSA interface was developed and it estimates parameters related to the Arrhenius and Aquilanti-Mundim formulations from experimental data, thus drawing a parallel between the usual model  $Q_{10}$  and the model proposed here  $Q_{d10}$ . *iii)* We also propose deep learning models that use artificial neural networks and learn from a spectral database to make inferences about structural properties of molecules and perform spectral reconstructions. The classifications performed by the ResNet34 model and the proposed tabular model show a good performance using infrared and mass spectra, with F1-Score values between 0.83 and 0.88 and accuracies of 96%. The reconstruction models showed a mean square error of 0.005 and 0.007 in the reconstruction of the mass and infrared spectra, respectively. Thus, it is concluded that the development of these models suggests that neural networks are suitable for the search for new spectral analysis techniques and represent a step towards the creation of autonomous analysis techniques. Finally, the codes developed here present very intuitive and user-friendly interfaces in order to disseminate the proposed new concepts and the authors hope that they can also be used as powerful research tools and didactic tools for teaching thermodynamics, process kinetics and related issues.

**KEYWORDS:** deformed Arrhenius theory, chemical kinetics,  $Q_{10}$  temperature coefficient, biological processes, machine learning, software development.

## LISTA DE FIGURAS

- Figura 1:** Fluxograma dos temas, áreas, aplicações e publicações deste trabalho....3
- Figura 2:** Fluxograma básico para construção do algoritmo GSA. Como discutido previamente, o GSA é um método de ML estocástico que resolve problemas de otimização e é uma técnica probabilística para encontrar o mínimo global ótimo de uma dada função custo. ....16
- Figura 3:** Representação genérica do gráfico da transitividade  $\gamma\beta = 1/Ea$  vs.  $\beta = 1RT$ . O comportamento *Arrhenius* é dado pela linha amarela paralela ao eixo  $\beta$  e os desvios sub e super *Arrhenius* são representados nas linhas azuis. ....17
- Figura 4:** Gráficos de *Arrhenius* comparando as constantes cinéticas experimentais e ajustadas para: reação de tautomerização ceto-enólica (comportamento sub-*Arrhenius* sob tunelamento profundo), reação  $OH.+H_2 \rightarrow H_2O + H$ . (comportamento sub-*Arrhenius* sob regime de tunelamento moderado), transferência de hidreto com catálise enzimática (comportamento super-*Arrhenius*) e reação  $OH.+HBr \rightarrow Br.+H_2O$  (comportamento anti-*Arrhenius*). As referências dos dados experimentais serão apresentadas na Tabela 1.....19
- Figura 5:** Gráficos de *Arrhenius* (superior) e da transitividade (inferior) para a dependência da temperatura do tempo de relaxamento do carbonato de propileno. Os símbolos de diamante representam os valores de transitividade obtidos numericamente e suavizados com o filtro *Savitzky – Golay*. As linhas vermelhas enfatizam duas regiões onde a dependência da transitividade com a temperatura é linearizada, conforme esperado pela lei de *Aquilanti-Mundim*. ....22
- Figura 6:** Gráfico de *Arrhenius* obtido a partir do código *Transitivity* para a reação  $OH.+HCl \rightarrow Cl.+H_2O$  utilizando a TST com as correções de Bell35, Bell58, ST e para *d*-TST. Os dados experimentais de Ravishankara e colaboradores<sup>104</sup> estão disponíveis para comparação e foram exibidos como pontos inteiros .....24
- Figura 7:** Na parte superior estão os gráficos de *Arrhenius* obtidos pelo código *Transitivity* para a reação  $NH_3 + OH. \rightarrow NH_2.+H_2O$  utilizando as formulações de *Kramer* e *Collins-Kimball*. Na parte inferior estão os gráficos do coeficiente de transmissão de *Kramer* e da constante do limite de difusão de *Smoluchowski* como uma função inversa da temperatura.....25
- Figura 8:** Reagentes, intermediários e estruturas de transição e todas as etapas elementares da formação de chalcona 4 catalisada por base a partir de benzaldeído

1 e acetofenona 2: (R1) primeira enolização, (R2) formação da ligação C-C, (R3) equilíbrio prótico, (R4) segunda enolização, e (R5) eliminação de hidróxido e formação da ligação C=C. A nomenclatura e indexação utilizada no trabalho de Chang e Perrin<sup>117</sup> foram mantidas para facilitar comparações.....27

**Figura 9:** Diagrama de energia livre de Gibbs para a formação de chalcona catalisada por base, experimentalmente por Chang e Perrin<sup>117</sup>, e calculadas em ambientes aquoso e acetonitrila nos níveis de teoria B2PLYP/AUG-cc-pVDZ//B3LYP/6-31+G(3df,3dp). Nomenclatura das estruturas foram definidas na Figura 8.....30

**Figura 10:** Diagrama de energia livre de Gibbs para as etapas R4 e R5 enfatizando os efeitos: isotópicos, de temperatura e de solvente contínuo.....34

**Figura 11:** Coeficiente de temperatura  $Q_{d10}$  generalizado. (A) Comportamento do coeficiente  $Q_{d10}$  como uma função do parâmetro de deformação  $d$ , assumindo a energia de ativação  $E_0$  como uma constante. Em casos extremos nos quais  $d = 0$ ,  $kT + 10kT \cong 2$  e a o modelo padrão é recuperado. (B) Comportamento do coeficiente  $Q_{d10}$  de acordo com o parâmetro de deformação  $d$  e a energia de ativação  $E_0$ .....38

**Figura 12:** Efeito da temperatura no metabolismo do zooplâncton do ártico pelo modelo de coeficiente deformado ( $Q_{d10}$ ). (A) Gráficos de Arrhenius deformados nos casos de respiração ( $C_R$ ) e taxas de excreção de amônia ( $N_E$ ) e fosfato ( $P_E$ ). Energia de ativação, usando d-Arrhenius para as taxas de (B) respiração ( $C_R$ ), (C) excreção de amônia ( $N_E$ ) e (D) excreção de fosfato ( $P_E$ ), em ( $C_R$ ,  $N_E$  e  $P_E$ )  $\mu mol CZOO - 1day - 1$ . Coeficiente deformado ( $Q_{d10}$ ) para as taxas de (E) respiração ( $C_R$ ), (F) excreção de amônia ( $N_E$ ) e (G) excreção de fosfato ( $P_E$ ).....41

**Figura 13:** Sensibilidade à temperatura de processos de decomposição do SOC através do modelo de coeficiente deformado ( $Q_{d10}$ ). Influência da temperatura na taxa de decomposição do SOC em quatro sistemas diferentes: TL1: camada superficial do solo do arrozal – local 1; TL2: camada superior do solo do planalto - local 2; SL1: camada de subsolo do arrozal – local 1; e SL2: camada de subsolo do planalto – local 2. (A) Taxas de decomposição experimental (pontos) com o aumento da temperatura e o ajuste (linhas sólidas) usando o modelo apresentado na Equação (12); (B) coeficientes  $Q_{d10}$  versus temperatura; (C) comportamento da energia de ativação com o aumento da temperatura. ....43



**Figura 14:** Influência da temperatura na taxa de decomposição do SOC em quatro sistemas diferentes: superfície de floresta estacional decidual, superfície de pastagem semiárida, floresta perene e floresta tropical. (A) Taxas de decomposição experimental (pontos) com o aumento da temperatura e o ajuste (linhas sólidas) usando o modelo apresentado na Equação (12); (B) coeficientes  $Q_{d10}$  versus temperatura; (C) variação da energia de ativação com a temperatura. ....44

**Figura 15:** Sensibilidade à temperatura nas reações fisiológicas dos sistemas respiratório e cardiovascular através do modelo de coeficiente deformado ( $Q_{d10}$ ). (A) Os resultados experimentais e o modelo proposto exibiram um comportamento convexo do gráfico  $\ln(k) \times 1/T$  o que indicou um desvio da teoria de Arrhenius padrão, indicado por uma linha tracejada. Os resultados experimentais são indicados por pontos sólidos e o modelo proposto é indicado por uma linha sólida; (B) em ambos os casos, os coeficientes  $Q_{d10}$  produziram uma razão de taxas de  $k_d(T2)/k_d(T1) \approx 2$ ; (C) a energia de ativação muda abruptamente com a temperatura; (D) os processos analisados foram, neste caso, do tipo super-Arrhenius ( $d > 0$ ). ....45

**Figura 16:** Convergência na resposta à temperatura da respiração foliar entre biomas e tipos funcionais de plantas por meio do modelo de coeficiente deformado ( $Q_{d10}$ ). Gráfico de Arrhenius deformado das taxas de respiração das plantas para biomas florestais temperados e tipos funcionais de plantas, incluindo plantas decíduas temperadas de folha larga. (A) Gráfico de Arrhenius deformado, (B) coeficiente  $Q_{d10}$  e (C) energia de ativação deformada. ....47

**Figura 17:** Gráfico de Arrhenius deformado para taxa de consumo de oxigênio das mitocôndrias do fungo eurythermal killifish *Fundulus heteroclitus*: (A) dados experimentais, gráfico de Arrhenius padrão (linha tracejada) e o modelo proposto (linha contínua). Os parâmetros de deformação obtidos por ajuste aos dados experimentais são  $d = 0.1982$ ,  $E_0 = 0.1029 \times 10^2 kJ/mol$  e  $A = 0.3380 \times 10^7$ ; (B) coeficiente  $Q_{d10}$ ; (C) energia de ativação deformada. ....48

**Figura 18:** Modelo preditivo genérico desenvolvido neste trabalho. ....50

**Figura 19:** Representação de um neurônio matemático.  $x_n$  são os sinais de entrada,  $w_n$  são os pesos, ( $u$ ) é o resultado da combinação linear,  $f(u)$  é a função de ativação e ( $y$ ) é o sinal de saída (*output*). ....53

**Figura 20:** Representações genéricas de RNA's diretas. ....54

<b>Figura 21:</b> Função Limiar.....	55
<b>Figura 22:</b> Comportamento das funções de ativação: linear, ReLU, sigmoide e tangente hiperbólica .....	56
<b>Figura 23:</b> A esquerda um espectro IR e a direita o mesmo espectro IR concatenado com o espectro MS. ....	63
<b>Figura 24:</b> Métricas de validação para os modelos CNN e tabular utilizando os diferentes conjuntos de dados espectrais: MS, IR e IR + MS. ....	66
<b>Figura 25:</b> Acurácia, recall, especificidade, precisão e F1 – Score para cada classe individualmente, para os modelos CNN e Tabular utilizando o conjunto de dados espectrais de IR. ....	67
<b>Figura 26:</b> Acurácia, recall, especificidade, precisão e F1 – Score para cada classe individualmente, para os modelos CNN e Tabular utilizando o conjunto de dados espectrais de IR + MS.....	68
<b>Figura 27:</b> Espectro IR da molécula bis (2-etoexietil) obtido no NIST e a reconstrução gerada pelo modelo.....	72
<b>Figura 28:</b> Espectro MS da molécula n-Decanal obtido no NIST e a reconstrução gerada pelo modelo.....	72
<b>Figura 29:</b> Passo a passo da instalação e configuração dos programas Transitivity e dQ <sub>10</sub> -GSA.....	99
<b>Figura 30:</b> Janela principal do código Q <sub>d10</sub> - GSA. Esta tela mostra os dados de entrada do usuário bem como as ferramentas disponíveis no software. 1: Parâmetros do GSA; 2: Dados de entrada; 3: Parâmetros do modelo; 4: Ferramentas para a exibição de gráficos; 5: Formatos de entrada para temperatura e constante cinética; 6: Botões de ação. ....	103
<b>Figura 31:</b> Exemplos dos gráficos que podem ser gerados através das opções contidas na sub janela “Plot”, após a realização do ajuste. No gráfico superior direito é possível visualizar os pontos que correspondem aos dados experimentais introduzidos no software, bem como as curvas dos modelos de Arrhenius e <i>d</i> -Arrhenius geradas através do ajuste.....	104

## LISTA DE TABELAS

<b>Tabela 1:</b> Parâmetros ajustados para as formulações AM, ASCC, NTS e VFT utilizando o código <i>Transitivity</i> para os processos de Tautomerização ceto-enólica <sup>35</sup> , reação $OH. + H_2$ <sup>90</sup> , catálise enzimática <sup>92</sup> e reação $OH. + Br$ <sup>98</sup> . Energy ( $E_a$ , $\epsilon \ddagger$ , $E_v$ e $E_0$ ) estão em <i>cal mol</i> e a temperatura ( $T_0$ e $B$ ) em Kelvin ( $K$ ). As unidades dos fatores pré-exponenciais podem ser obtidas em suas referências.....	21
<b>Tabela 2:</b> Frequências imaginárias ( $cm - 1$ ) e energias livre de Gibbs ( <i>kcal mol</i> - 1) as espécies da Figura 8 para diferentes solventes. ....	30
<b>Tabela 3:</b> Parâmetros usados no ajuste do modelo deformado de Arrhenius para os dados experimentais. ....	44
<b>Tabela 4:</b> Sistemas respiratório e cardiovascular Parâmetros usados no modelo deformado de <i>Arrhenius</i> como consequência do ajuste do $Q_{d10}$ a partir dos dados experimentais. ....	46
<b>Tabela 5:</b> Ao meio a lista das classes analisadas, dentre elas: grupos funcionais e fragmentos diversos. A direita a quantidade de compostos que apresentam o fragmento de referência e a esquerda o índice de cada classe. ....	63
<b>Tabela 6:</b> As moléculas “ácido cumarílico” e “ácido 3-nitro-4-clorobenzóico” fazem parte do nosso dataset de validação e foram avaliadas para os modelos CNN e Tabular treinados com os conjuntos de dados IR e IR + MS. As colunas “Rótulo” são vetores que indicam quais grupos funcionais a molécula apresenta (padrão ouro), enquanto as colunas “Predição” são os vetores de saída dos modelos que indicam quais grupos funcionais foram preditos para aquela molécula. Na parte de baixo da tabela estão expressas as métricas: F1-Score, acurácia, recall, precisão e especificidade calculadas diretamente entre dos vetores de predição e rótulo correspondentes.....	69
<b>Tabela 7:</b> As moléculas “m-hidroxidifenilamina” e “iodometano” fazem parte do nosso dataset de validação e foram avaliadas para os modelos CNN e Tabular treinados com os conjuntos de dados IR e IR + MS. As colunas “Rótulo” são vetores que indicam quais grupos funcionais a molécula apresenta (padrão ouro), enquanto as colunas “Predição” são os vetores de saída dos modelos que indicam quais grupos funcionais foram preditos para aquela molécula. Na parte de baixo da tabela estão expressas as métricas: F1-Score, acurácia, recall, precisão e especificidade calculadas diretamente entre dos vetores de predição e rótulo correspondentes.....	70

## LISTA DE SIGLAS E ABREVIATURAS

- AM – Aquilanti-Mundim
- ASCC – Aquilanti-Sanchez-Coutinho-Carvalho
- Acc – Acurácia
- BOMD – Dinâmica Molecular de Born-Oppenheimer (*Born-Oppenheimer Molecular Dynamic*)
- CNN – Rede Neural Convolutacional (*Convolutional Neural Network*)
- CPMD – Dinâmica Molecular de Car-Parrinello (*Car-Parrinello Molecular Dynamic*)
- CSA – Resfriamento Simulado Clássico (*Classical Simulated Annealing*)
- DFT – Teoria do Funcional da Densidade (*Density Functional Theory*)
- DL – Aprendizagem Profunda (*Deep Learning*)
- DNN – Rede Neural Profunda (*Deep Neural Network*)
- FN – Falso Negativo
- FP – Falso Positivo
- FSA – Resfriamento Simulado Rápido (*Fast Simulated Annealing*)
- GBL – Aprendizado Baseado em Gradiente (*Gradient-based Learning*)
- GPU – Unidade de Processamento Gráfico (*Graphics Processing Unit*)
- GSA – Resfriamento Simulado Generalizado (*Generalized Simulated Annealing*)
- IA – Inteligência Artificial
- IR – Infra Vermelho (*Infra Red*)
- IUPAC – União Internacional de Química Pura e Aplicada (*International Union of Pure and Applied Chemistry*)
- LMSC – Laboratório de Modelagem de Sistemas Complexos
- LSTM – Memória Longa de Curto Prazo (*Long short-term memory*)
- ML – Aprendizado de Máquina (*Machine Learning*)
- MS – Espectrometria de Massa (*Mass Spectrometry*)
- MTD – Meta Dinâmica
- NIST – Instituto Nacional de Ciência e Tecnologia dos Estados Unidos (*United States National Institute for Science and Technology*)
- NTS – *Nakamura-Takayanagi-Sato*
- PES – Superfícies de Energia Potencial (*Potential Energy Surface*)
- PIMD – Dinâmica Molecular de Integrais de Trajetória (*Path-Integral Molecular Dynamic*)

Prec – Precisão

QSAR – Relação de Atividade de Estrutura Quantitativa (*Quantitative Structure Activity Relationship*)

QTEA – Química Teórica e Estrutural de Anápolis

RCS – Etapa Lenta de Controle (*Rate Control Step*)

RNA – Rede Neural Artificial

Rec – Recordar (Recall)

SA – Resfriamento Simulado (*Simulated Annealing*)

SGD – Gradiente Descendente Estocástico (*Stochastic Gradient Descent*)

SOC – Carbono Orgânico do Solo (*Soil Organic Carbon*)

ST – *Skodje-Truhlar*

SVM – Máquinas de Vetor de Suporte (*Support Vector Machines*)

Spec – Especificidade

TN – Negativo Verdadeiro (*True Negative*)

TP – Positivo Verdadeiro (*True Positive*)

TSH – Salto de Superfície de Trajetória (*Trajectory Surface Hopping*)

TST – Teoria do Estado de Transição (*Transition State Theory*)

UEG – Universidade Estadual de Goiás

VFT – *Vogel-Fulcher-Tamman*

d-TST – Teoria do Estado de Transição Deformada (*Deformed Transition State Theory*)

## SUMÁRIO

<b>INTRODUÇÃO</b> .....	1
<b>CAPÍTULO 1</b> .....	4
<b>1. TRANSITIVITY</b> .....	4
1.1 Introdução .....	4
1.2 Constante Cinética de Processos Reacionais.....	6
1.2.1 Funções de Partição.....	7
1.2.2 Teoria do Estado de Transição deformada (d-TST) .....	8
1.2.3 Correções de Tunelamento .....	10
1.2.4 Efeitos de solvente na constante cinética da reação .....	11
1.3 Processo de Otimização via GSA .....	13
1.4 Energia de Ativação e o Gráfico da Transitividade .....	16
1.5 Exemplos e aplicações.....	18
1.5.1 Ajustes de função – Gráficos de <i>Arrhenius</i> .....	18
1.5.2 Ajustes de função – Gráfico da transitividade.....	22
1.5.3 Estimativa de constantes cinéticas de reação .....	23
1.6 Síntese de chalconas – uma aplicação do <i>Transitivity</i> .....	26
1.6.1 Procedimentos Computacionais .....	28
1.6.2 Influência do solvente no perfil da energia livre de <i>Gibbs</i> .....	29
1.6.3 Individualização de fatores físico-químicos na cinética da condensação de <i>Claisen-Schmidt</i> .....	31
1.6.4 Relação com o trabalho de Chang e Perrin .....	34
<b>CAPÍTULO 2</b> .....	36
<b>2. Q<sub>d10</sub> – GSA</b> .....	36
2.1 Introdução .....	36

2.2	Coeficiente da temperatura $Q_{10}$ .....	37
2.3	Exemplos e aplicações.....	39
2.3.1	Metabolismo do Zooplâncton do Ártico.....	39
2.3.2	Processos de Decomposição de Carbono Orgânico do Solo .....	42
2.3.3	Processos de Decomposição de SOC em diferentes tipos de solo .....	44
2.3.4	Reações Fisiológicas dos Sistema Respiratório e Cardiovascular .....	45
2.3.5	Respiração Foliar entre Biomas e tipos Funcionais de Plantas .....	46
2.3.6	Consumo de Oxigênio nas Mitocôndrias do Fungo <i>Heteroclito Eurythermal Killifish</i> .....	47
<b>CAPÍTULO 3 .....</b>		<b>49</b>
<b>3. DEEP LEARNING NA ANÁLISE DE TÉCNICAS ESPECTROSCÓPICAS .....</b>		<b>49</b>
3.1	Introdução .....	49
3.2	Bases Teóricas e metodologias .....	50
3.2.1	Inteligência artificial .....	50
3.2.2	O Neurônio Matemático .....	52
3.2.3	Redes Neurais Artificiais.....	53
3.2.4	Redes Neurais Convolucionais.....	59
3.2.5	Codificação de moléculas para uso em ML e DL.....	60
3.3	Caracterização de grupos funcionais moleculares .....	62
3.3.1	Treinamento e hiperparametrização .....	63
3.3.2	Validação .....	64
3.4	Espectros moleculares gerados a partir de descritores moleculares .....	71
<b>CONSIDERAÇÕES FINAIS E PERSPECTIVAS .....</b>		<b>73</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>		<b>77</b>

<b>APÊNDICE A – PUBLICAÇÕES</b> .....	94
A.1 – “Transitivity”: A code for Computing Kinetic and Related Parameters in Chemical Transformations and Transport Phenomena .....	94
A.2 – Temperature coefficient ( $Q_{10}$ ) and its applications in biological systems: Beyond the Arrhenius theory .....	95
A.3 – Topography of the free energy landscape of Claisen-Schmidt condensation: solvent and temperature effects on the rate-controlling step .....	96
A.4 – Reply to the ‘Comment on “Topography of the Free Energy Landscape on the Claisen–Schmidt Condensation: Solvent and Temperature Effect in the Rate-Controlling Step” by N. D. Coutinho, H. G. Machado, V. H. Carvalho-Silva and W. A. da Silva, Phys. Chem. Chem. Phys., 2021, 23, 6738’ .....	97
<b>APÊNDICE B – LIDANDO COM AS INTERFACES GRÁFICAS DOS CÓDIGOS E MODELOS DESENVOLVIDOS</b> .....	98
B.1 - Download, Instalação e Execução dos Softwares.....	98
B.2 - Lidando com o código Transitivity .....	99
B.2.1 - Cinética e Parâmetros Relacionados.....	100
B.2.2 - Ajuste de Otimização GSA .....	101
B.2.3 - Dinâmica Molecular .....	101
B.3 - Lidando com o código Qd10 – GSA.....	102



## INTRODUÇÃO

As primeiras investigações acerca de uma Inteligência Artificial (IA) datam de 1950 com o matemático Alan Turing<sup>1</sup>. Desde então o tema foi tratado como um dos maiores potenciais futurísticos da evolução da tecnologia, mas o custo computacional elevado e a escassez de dados foram determinantes na lenta evolução desta área. Os desafios mais recentes desta área envolvem a resolução de problemas com soluções intuitivas<sup>2</sup>, em outras palavras, tarefas que são realizadas facilmente por pessoas, mas que são difíceis de serem descritas formalmente, como por exemplo o reconhecimento de imagens e textos.

Dentre as diversas abordagens da IA, o Aprendizado de Máquina (ML, do inglês: *Machine Learning*) é a área com maior destaque. Em suma, algoritmos de ML extraem informações de dados brutos e os representa através de um modelo matemático. Posteriormente este modelo é utilizado para fazer inferências sobre os dados analisados e, principalmente, acerca de novos dados. O diferencial do ML é seu modelo de aprendizagem que aplica análises estatísticas sobre um conjunto de dados, ao invés de programar regras e rotinas específicas para resolução de um problema.

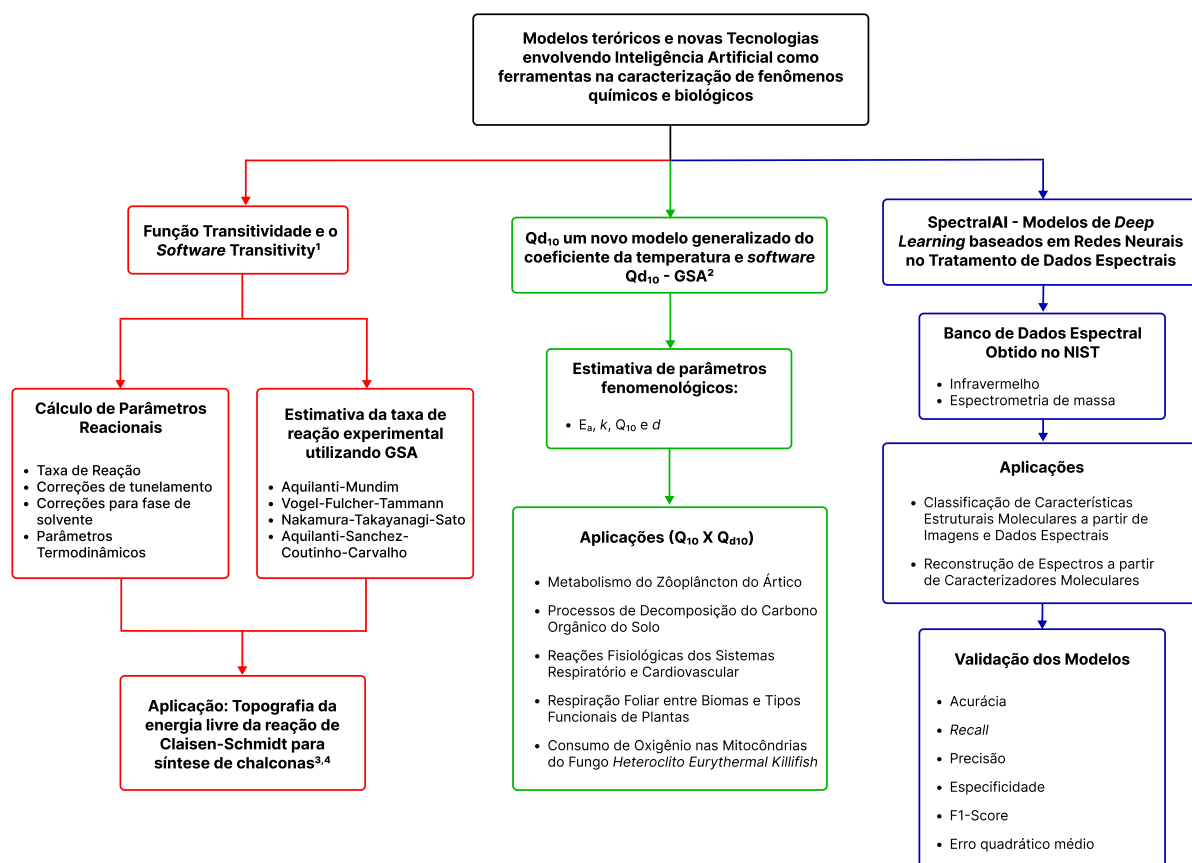
Modelos de aprendizagem estatística baseados em ML já vêm sendo utilizados há décadas na química, no campo da química computacional, por exemplo: na modelagem química “Relação Quantitativa entre Estrutura e Atividade” (QSAR, do inglês: *Quantitative Structure Activity Relationship*)<sup>3-6</sup>; na estimativa de parâmetros fenomenológicos<sup>7</sup> e na descrição de Superfícies de Energia Potencial (PES, do inglês: *Potential Energy Surface*)<sup>8</sup>. Recentemente, com o advento de métodos sofisticados de Aprendizagem Profunda (DL, do inglês: *Deep Learning*), estes tipos de análises baseadas em dados têm se tornado rotina em diversas aplicações químicas e biológicas, como no desenvolvimento de drogas<sup>5,9-11</sup> na predição de propriedades químicas<sup>12-15</sup> e até em cálculos quânticos<sup>16-19</sup>.

Nas últimas décadas nossos grupos de pesquisa: Laboratório de Modelagem de Sistemas Complexos (LMSC), da Universidade de Brasília (UnB); e Química Teórica e Estrutural de Anápolis (QTEA) da Universidade Estadual de Goiás (UEG); têm se dedicado ao estudo cinético e mecanístico de processos reativos e não

reativos<sup>20-29</sup>, no desenvolvimento de novas teorias e metodologias de análise para realização destes estudos<sup>7,30</sup>, na produção de algoritmos de ML<sup>7</sup> e softwares nas áreas de cinética<sup>31,32</sup> e caracterização espectroscópica<sup>33</sup>. Neste sentido, vem se tornando cada vez mais necessário o desenvolvimento de tecnologias com plataformas que sistematizem essas técnicas e sirvam como ferramenta para os pesquisadores. Com isso, aqui nós desenvolvemos novos formalismos e modelos matemáticos nas áreas de cinética e espectroscopia, além de técnicas de análise utilizando modelos de ML e DL que serão apresentados juntamente com *softwares* desenvolvidos para aplicação destes (ver publicações no apêndice A). Este trabalho está dividido e será apresentado em três capítulos (ver Figura 1):

- Capítulo 1: A função transitividade e o *software Transitivity* no estudo da cinética de mecanismos reacionais e outros processos.
- Capítulo 2: O modelo matemático  $Q_{d10}$  e sua aplicação em processos biológicos utilizando o *software  $Q_{d10}$ -GSA*.
- Capítulo 3: Modelos de DL para o tratamento e análise de dados espectrais de infravermelho e espectrometria de massa e a interface *web SpectralAI*.

Além do desenvolvimento dos novos conceitos, formalismos e modelos matemáticos, este trabalho também objetivou o desenvolvimento de *softwares* com interfaces gráficas intuitivas e amigáveis no intuito de difundir, incentivar e facilitar o uso dos conceitos propostos.



<sup>1</sup>H. G. Machado, F. O. Sanches-Neto, N. D. Coutinho, K. C. Mundim, F. Palazzetti, and V. H. Carvalho-Silva, "Transitivity: A code for computing kinetic and related parameters in chemical transformations and transport phenomena" *Molecules*, vol. 24, no. 19, 2019, doi: 10.3390/molecules24193478.

<sup>2</sup>K. C. Mundim, S. Baraldi, H. G. Machado, and F. M. C. Vieira, "Temperature coefficient (Q10) and its applications in biological systems: Beyond the Arrhenius theory" *Ecological Modelling*, vol. 431, no. December, 2020, doi: 10.1016/j.ecolmodel.2020.109127.

<sup>3</sup>N. D. Coutinho, H. G. Machado, V. H. Carvalho-Silva, and W. A. da Silva, "Topography of the free energy landscape of Claisen-Schmidt condensation: solvent and temperature effects on the rate-controlling step" *Phys. Chem. Chem. Phys.*, p., 2021, doi: 10.1039/D0CP05659F.

<sup>4</sup>N. Dantas Coutinho, H. Gontijo Machado, H. Carvalho-Silva, W. Alves da Silva, and V. Henrique Carvalho, "Reply to 'Comment on "Topography of the Free Energy Landscape on the Claisen-Schmidt Condensation: Solvent and Temperature Effect in the Rate-Controlling Step"'" *Phys. Chem. Chem. Phys.*, 2021, doi: 10.1039/D1CP02799A.

**Figura 1:** Fluxograma dos temas, áreas, aplicações e publicações deste trabalho.

# CAPÍTULO 1

## TRANSITIVITY

### 1.1 Introdução

Para o tratamento adequado de processos reativos e não reativos (transformações químicas e fenômenos de transporte) e para a compreensão de mecanismos complexos, a análise cinética é de grande importância, principalmente se há dependência da temperatura em processos que envolvem a taxa de reação,  $k(T)$ , e frequentemente a baixas temperaturas onde desvios da linearidade do *plot* de *Arrhenius* são observados. A fenomenologia do modelo de *Arrhenius* abrange uma gama de processos que são majoritariamente governados por regimes energéticos e de colisão, principalmente a temperaturas altas e moderadas. De uma lista que é constantemente atualizada, nos referimos aqui à alguns casos selecionados, como química de combustão<sup>34</sup>, fase condensada<sup>35</sup>, reações atmosféricas e astroquímicas<sup>36,37</sup>, processos envolvendo a preservação e envelhecimento de alimentos e medicamentos<sup>38,39</sup>, bem como geoquímica básica<sup>40</sup> e ambientes bioquímicos<sup>41,42</sup>.

Entretanto, a baixas temperaturas fatores energéticos passam a ter uma menor contribuição na dinâmica do processo e regimes que antes não eram considerados passam a ter maior influência e podem modificar a dinâmica e alterar sensivelmente a velocidade destes processos. A existência destes outros regimes é sugerida ao analisar casos de desvios na lei de *Arrhenius* e é classificada da seguinte maneira: (i) cinética super-*Arrhenius*, curvatura convexa no *plot* de *Arrhenius*, na qual o fenômeno basicamente é cessado a baixas temperaturas; (ii) cinética sub-*Arrhenius*, curvatura côncava no *plot* de *Arrhenius*, na qual mecanismos reativos de tunelamento quântico são propícios a baixas temperaturas; (iii) cinética anti-*Arrhenius*, energia de ativação aparentemente negativa, na qual geralmente processos são limitados por requerimentos estereodinâmicos.

Para entender melhor estes processos e os regimes envolvidos, neste trabalho propomos a função transitividade:  $\gamma(T)$ ; uma nova perspectiva sobre o gráfico de *Arrhenius* que permite linearizar os desvios na lei de *Arrhenius* e assim

classificar e interpretar estas leis empíricas de forma a obter interpretações microscópica dos parâmetros fenomenológicos.

Para estes casos (sejam experimentais ou teóricos) cujos processos que envolvem a constante cinética são difíceis de serem interpretados (por complexidades moleculares, condições extremas ou impossibilidade de solução exata para equação de *Schrödinger*), o uso de abordagens teóricas fenomenológicas e semiclássicas que vão além do modelo de *Arrhenius* são de utilidade crescente e nesse sentido algumas leis empíricas foram propostas, por exemplo: e.g., *Kooij*<sup>43</sup>, lei das potências<sup>44</sup>, *Vogel-Fulcher-Tammann* (VFT)<sup>44-46</sup>, *Nakamura-Takayanagi-Sato* (NTS)<sup>47</sup>, *Aquilanti-Mundim* (AM)<sup>30</sup> e *Aquilanti-Sanchez-Coutinho-Carvalho* (ASCC)<sup>27</sup>. Além disso, para compreender e prever uma variedade de processos cinéticos de alta complexidade molecular e desvios da lei de *Arrhenius*, a Teoria do Estado de Transição (TST, do inglês: *Transition State Theory*) continua sendo uma excelente formulação. Neste sentido, ferramentas de software para calcular e interpretar as constantes cinéticas são bastante úteis em pesquisas de geoquímica, bioquímica e química de materiais, pois permitem grandes avanços nas análises apenas utilizando técnicas computacionais. Uma ampla gama de códigos tem sido desenvolvida no intuito de calcular constantes cinéticas em fase gasosa, líquida e em sólidos utilizando a TST e suas variantes: *Polyrate*<sup>48</sup>, *TheRate*<sup>49</sup>, *MultiWell*<sup>50</sup>, *TAMkin*<sup>51</sup>, *Mesmer*<sup>52</sup>, *RMG*<sup>53</sup>, *APUAMA*<sup>54</sup>, *KiSThelP*<sup>55</sup>, *FRIGUS*<sup>56</sup> e *Eyringpy*<sup>57</sup> são excelentes opções disponíveis para trabalhar com cinética de reações químicas.

Visando contribuir para uma melhor compreensão e caracterização desses fenômenos, desenvolvemos um novo código, denominado *Transitivity*, com uma interface gráfica amigável capaz de executar diversos procedimentos para calcular parâmetros cinéticos e outros parâmetros relacionados em reações químicas e fenômenos de transporte, além de fornecer ferramentas para cálculo e avaliação da função transividade aqui proposta.

Ao decorrer deste Capítulo 1 será descrito, detalhadamente, todos os conceitos básicos relativos a constante cinética de processos reacionais, formando a base dos fenômenos estudados neste capítulo e dos procedimentos de cálculos implementados no *Transitivity*: desde o cálculo da constante cinética utilizando a TST e suas correções; até o uso do “Resfriamento Simulado Generalizado” (GSA, do

inglês: *Generalized Simulated Annealing*) para, a partir de dados obtidos experimentalmente, realizar inferências fenomenológicas utilizando um compilado de teorias cinéticas. Paralelamente, será apresentada uma série de aplicações em casos conhecidos da literatura de modo a validar a metodologia implementada no código, além de abordar separadamente um estudo acerca do perfil de energia livre na cinética da síntese de chalconas pelo mecanismo de *Claisen-Schmidt*. Será mostrado também que o gráfico da transitividade é uma ferramenta útil para linearizar curvaturas não-*Arrhenius*, permitindo assim identificar mudanças de comportamento durante o processo que não são percebidas no gráfico de *Arrhenius*.

Ao final do texto, a seção de apêndice B fornece instruções relativas à obtenção, instalação e dicas de utilização do código desenvolvido.

## 1.2 Constante Cinética de Processos Reacionais

A Teoria do Estado de Transição, TST, é a ferramenta mais popular usada para estudar a cinética de reações químicas com um complexo ativo definido. Como de costume, o símbolo ( $\ddagger$ ) denota propriedades pertencentes ao complexo do estado de transição. Para uma reação bimolecular geral, do tipo  $R_1 + R_2 \rightarrow TS^\ddagger \rightarrow \text{Produtos}$ , é necessário calcular as funções de partição  $Q_1$ ,  $Q_2$  e  $Q^\ddagger$  referentes aos reagentes  $R_1$  e  $R_2$  e ao estado de transição, respectivamente. A uma dada temperatura absoluta  $T$ , a constante cinética da reação é dada por:

$$k^{\text{TST}}(T) = \frac{k_B T}{h} \frac{Q^\ddagger}{Q_1 Q_2} e^{-\left(\frac{\varepsilon^\ddagger}{RT}\right)} \quad (1)$$

na qual  $h$  é a constante de *Planck* [ $h \cong 6,626 * 10^{-34} \text{ (J * s)}$ ];  $k_B$  é a constante de *Boltzmann* [ $k_B \cong 1,38 * 10^{-23} \text{ (J/K)}$ ];  $R$  é a constante dos gases ideais [ $R \cong 8,31 \text{ (J/mol K)}$ ];  $T$  é a temperatura absoluta em Kelvin ( $K$ );  $Q_1$ ,  $Q_2$  e  $Q^\ddagger$  são as funções de partição para os reagentes 1, 2 e para o estado de transição, respectivamente; e  $\varepsilon^\ddagger$  é a altura efetiva da barreira de energia da reação (eventualmente adicionada à correção da energia de ponto zero). A constante dos gases ideais  $R$  pode ser descrita como  $R = k_B * N_a$ , onde  $N_a \cong 6,02 * 10^{23} \text{ moleculas}$  é o número de *Avogadro*;

A Equação (1) é uma reformulação da clássica equação de *Arrhenius*<sup>58</sup>, expressa pela Equação (2):

$$k(T) = A e^{-\left(\frac{E_0}{RT}\right)} \quad (2)$$

em que  $A$  é o fator pré-exponencial; e  $E_0$  é a altura de barreira da reação.

Esta formulação de *Arrhenius* data de 1889<sup>58</sup>, mas só em 1916 Max Trautz propôs a teoria cinética de colisões<sup>59</sup>, interpretando o fator pré-exponencial  $A$  como sendo um fator de colisão, baseado na ideia de que, além de ter a energia necessária, é necessário que as moléculas entrem em colisão para que uma reação ocorra. Em 1935 Eyring propõe a existência do complexo ativado<sup>60</sup>, uma entidade intermediária do processo de reação. Eyring interpretou a altura de barreira energética de uma reação como uma diferença de energia entre os reagentes e este complexo ativado, dando origem à TST, descrita na Equação (1), onde o fator pré-exponencial  $A$  (fator de colisões) é expresso em função das funções de partição.

### 1.2.1 Funções de Partição

A formulação da TST, Equação (1), é descrita em termos das funções de partição  $Q$ , contabilizadas para cada entidade molecular considerando as contribuições: translacional; rotacional; vibracional; e eletrônica, de uma molécula, conforme a Equação (3).

$$Q = Q_{trans} Q_{rot} Q_{vib} Q_{elet} \quad (3)$$

Na Equação (3),  $Q_{trans}$  é a contribuição translacional tridimensional, expressa na Equação (4):

$$Q_{trans} = \frac{V}{\Lambda^3}; \quad \Lambda = h \left( \frac{1}{2\pi m k_B T} \right)^{1/2} \quad (4)$$

onde  $V$  é o volume; e  $m$  é a massa. O termo  $Q_{rot}$  é a contribuição rotacional, seu valor é  $Q_{rot} = 1$  para átomos e pode ser descrita na Equação (5) para moléculas lineares:

$$Q_{rot} = \frac{T}{\sigma \theta_{rot}}; \quad \theta_{rot} = \frac{hcB}{k_B} = \frac{h^2}{8\pi^2 k_B I}, \quad (5)$$

onde  $\sigma$  é o número de simetria;  $\theta_{rot}$  é a temperatura rotacional; que pode ser escrita em termos da constante rotacional  $B$  e da velocidade da luz  $c$ ; ou do momento de inércia  $I$ . Enquanto moléculas lineares apresentam somente um momento de inércia ( $I$ ), moléculas não-lineares apresentam três momentos de inércia ( $I_A$ ,  $I_B$  e  $I_C$ ) e, conseqüentemente, três temperaturas rotacionais ( $\theta_{rot}^A$ ,  $\theta_{rot}^B$  e  $\theta_{rot}^C$ ) e três constantes rotacionais ( $B_A$ ,  $B_B$  e  $B_C$ ). A função de partição rotacional  $Q_{rot}$  para moléculas não lineares pode ser descrita na Equação (6) em termos das

temperaturas rotacionais, na Equação (7) em termos das constantes rotacionais e na Equação (8) em termos dos momentos de inércia.

$$Q_{rot} = \frac{\sqrt{\pi}}{\sigma} \left[ \frac{T^{3/2}}{(\theta_{rot}^A \theta_{rot}^B \theta_{rot}^C)^{1/2}} \right] \quad (6)$$

$$Q_{rot} = \frac{\sqrt{\pi}}{\sigma} \left( \frac{k_B T}{hc} \right)^{3/2} \left( \frac{1}{B_A B_B B_C} \right)^{1/2} \quad (7)$$

$$Q_{rot} = \frac{\sqrt{\pi}}{\sigma} \left( \frac{8\pi^2 k_B T}{h^2} \right)^{3/2} (I_A I_B I_C)^{1/2} \quad (8)$$

O termo  $Q_{vib}$  é a contribuição vibracional, seu valor para átomos é  $Q_{vib} = 1$ , para moléculas poliatômicas é descrita pela Equação (9):

$$Q_{vib} = \prod_i \left( \frac{e^{-\frac{h\nu_i}{2k_B T}}}{1 - e^{-\frac{h\nu_i}{k_B T}}} \right) \quad (9)$$

onde  $i$  é o número de estados vibracionais; e  $\nu_i$  é a frequência vibracional do estado  $i$ . Por fim,  $Q_{elet}$  é a contribuição eletrônica, expressa na Equação (10):

$$Q_{elet} = g_E \quad (10)$$

onde  $g_E$  é a degenerescência para a excitação do estado fundamental.

### 1.2.2 Teoria do Estado de Transição deformada (d-TST)

A formulação da teoria do estado de transição deformada ( $d$ -TST) é dada pela Equação (11).

$$k_d(T) = \frac{k_B T}{h} \frac{Q^\ddagger}{Q_1 Q_2} \left( 1 - d \frac{\varepsilon^\ddagger}{k_B T} \right)^{1/d} \quad (11)$$

A nomenclatura da Equação (11) é a mesma usada na Equação (1). A formulação expressa na Equação (11) é derivada da lei de *Aquilanti-Mundim*<sup>30</sup> que pode ser expressa em sua forma clássica na Equação (12).

$$k_d(T) = A \left( 1 - d \frac{E_0}{RT} \right)^{1/d} \quad (12)$$

Uma opção alternativa e popular na literatura é usar a correção de tunelamento de Wigner, no entanto, em nossos estudos anteriores<sup>61,62</sup>, mostramos que  $d$ -TST é uma aproximação mais satisfatória também em vista do comportamento uniforme em toda a altura de barreira de energia.



### **Fenomenologia da d-TST**

No cálculo diferencial integral uma das maneiras de definir o limite fundamental exponencial<sup>63</sup> é expressá-lo segundo a Equação (13).

$$\lim_{x \rightarrow 0} (1 + xa)^{\frac{1}{x}} = e^a \quad (13)$$

A formulação de *Aquilanti-Mundim* é a aplicação do conceito do limite fundamental exponencial na equação de *Arrhenius*, deste modo surgindo o parâmetro  $d$ , na Equação (14):

$$\lim_{d \rightarrow 0} A \left( 1 - d \frac{E_0}{RT} \right)^{1/d} = A e^{-\left(\frac{E_0}{RT}\right)} \quad (14)$$

para  $d \rightarrow 0$  a equação clássica de *Arrhenius* é recuperada e o *plot* de *Arrhenius* ( $\ln k \times 1/T$ ) segue linear. Valores com  $d \neq 0$  promovem curvaturas no *plot* de *Arrhenius* e para classificar estes desvios utiliza-se as nomenclaturas super-*Arrhenius* ( $d > 0$ ) e sub-*Arrhenius* ( $d < 0$ ).

Casos super-*Arrhenius* ( $d > 0$ ) estão geralmente associados a fenômenos clássicos que a baixas temperaturas apresentam quedas mais acentuadas na velocidade do processo do que previsto pela lei de *Arrhenius*, podendo até haver uma “temperatura de morte” (onde o processo para de ocorrer). Alguns fenômenos ocorrem: na degradação de alimentos<sup>64</sup>, na catálise enzimática<sup>65</sup> e na difusão de substâncias em sistemas super-resfriados próximos à temperatura de transição vítrea<sup>66,67</sup>. Estes casos apresentam questões principalmente de complexidade estérica (enzimas, moléculas biológicas) que a partir de uma temperatura suficientemente baixa passam a governar a cinética dos processos em detrimento do fator energético, principal regulador desta cinética em temperaturas mais altas.

Casos sub-*Arrhenius* ( $d < 0$ ) estão geralmente associados a fenômenos quânticos nos quais, a baixas temperaturas, as partículas têm mais tempo para promover rotações internas e, por questões orientacionais, acabam facilitando a ocorrência do processo apesar das baixas frequências de colisão e fraca energia de interação. O tunelamento quântico, por exemplo, é um processo favorecido a baixas temperaturas, uma vez que a temperaturas moderadas e altas as partículas já têm energia suficiente para sobrepor a barreira. No intuito de entender o parâmetro de deformação ( $d$ ) um estudo anterior realizado por nosso grupo de pesquisa<sup>68</sup> sugere um modelo que descreve este parâmetro, expresso na Equação (15), para casos sub-*Arrhenius*:

$$d = -\frac{1}{3} \left( \frac{h\nu^\ddagger}{2\varepsilon^\ddagger} \right)^2 \quad (15)$$

na qual  $\varepsilon^\ddagger$  é a altura da barreira e  $\nu^\ddagger$  a frequência do complexo de transição ou frequência para cruzar a barreira, que cobre uniformemente a faixa do regime de tunelamento clássico ao moderado, mas precisa de um ajuste em casos raros de tunelamento profundo. Neste modelo, quanto maior for a altura de barreira e menor a frequência para a atravessar a barreira, menor é o valor do parâmetro  $d$ . Quando este parâmetro  $d$  tende a zero o efeito de tunelamento é negligenciado e o modelo de *Arrhenius* é recuperado. Importante ressaltar que o formalismo por trás deste modelo não se preocupa com os impactos na termodinâmica de *Tsallis*<sup>69</sup>, somente na sua descrição fenomenológica em casos sub-*Arrhenius*, de forma que o parâmetro  $d$  será sempre negativo (caso sub-*Arrhenius*). Para uma completa descrição do significado físico e fenomenológico do parâmetro  $d$  mais investigação ainda são necessárias.

### 1.2.3 Correções de Tunelamento

#### *Bell35 e Bell58*

Para cobrir a transição do tunelamento moderado ao profundo (adjetivos associados aos desvios da linearidade no gráfico de *Arrhenius* em casos sub-*Arrhenius*) em reações exotérmicas nós aplicamos ambas as correções de Bell35<sup>70,71</sup>, Equação (16):

$$\kappa_{Bell35} = \frac{\left[ \frac{1}{h\nu^\ddagger} - \frac{1}{k_B T} e^{\left( \frac{\varepsilon^\ddagger}{k_B T} - \frac{\varepsilon^\ddagger}{h\nu^\ddagger} \right)} \right]}{\frac{1}{h\nu^\ddagger} - \frac{1}{k_B T}} \quad (16)$$

e correção de Bell58<sup>71,72</sup> truncada no segundo termo (2T), Equação (17):

$$\kappa_{Bell58-2T} = \frac{\left( \frac{h\nu^\ddagger}{2k_B T} \right)}{\text{sen} \left( \frac{h\nu^\ddagger}{2k_B T} \right)} - \frac{e^{\left( \frac{\varepsilon^\ddagger}{k_B T} - \frac{\varepsilon^\ddagger}{h\nu^\ddagger} \right)}}{\left( \frac{k_B T}{h\nu^\ddagger} - 1 \right)} \quad (17)$$

que, embora não seja uniforme em toda a transição entre os regimes de tunelamento negligenciável e moderados, foram encontrados para se comportar suavemente o suficiente para executar adequadamente praticamente em todo o intervalo.

No entanto, esta formulação apresenta divergência na temperatura de cruzamento  $T_c = hv^\ddagger/\pi k_B$ . Os regimes de tunelamento podem ser delimitados dentro de quatro faixas de temperatura – negligenciável ( $T > 2T_c$ ), pequeno ( $2T_c > T > T_c$ ), moderado ( $T_c > T > T_c/2$ ) e profundo ( $T > T_c/2$ ) [60, 66, 67].

### Skodje e Truhlar (ST)

Para evitar divergências espúrias em  $T_c$  na formulação Bell58, em 1981, Skodje e Truhlar<sup>75</sup> fizeram uma generalização estendendo o tratamento da barreira parabólica, em sua aproximação para correção de tunelamento nas Equações (18) e (19):

$$\kappa_{ST} = \frac{\left(\frac{hv^\ddagger}{2k_B T}\right)}{\text{sen}\left(\frac{hv^\ddagger}{2k_B T}\right)} - \frac{e^{\left[\left(\frac{1}{k_B T} - \frac{1}{hv^\ddagger}\right)(\varepsilon^\ddagger - \Delta H)\right]}}{\left(\frac{k_B T}{hv^\ddagger} - 1\right)}, \quad \beta \leq \mathbb{C}, \quad (18)$$

$$\kappa_{ST} = \frac{1}{\left(\frac{k_B T}{hv^\ddagger} - 1\right)} \left\{ e^{\left[\left(\frac{1}{k_B T} - \frac{1}{hv^\ddagger}\right)(\varepsilon^\ddagger - \Delta H)\right]} - 1 \right\}, \quad \mathbb{C} \leq \beta \quad (19)$$

onde  $\beta = 1/RT$ ;  $\mathbb{C} = 1/k_B T_c = \pi/hv^\ddagger$  e  $\Delta H$  é a entalpia da reação.

### 1.2.4 Efeitos de solvente na constante cinética da reação

Duas abordagens acerca do efeito de solvente na constante cinética da reação estão implementadas e podem ser contabilizadas pelo *Transitivity*, são elas: a formulação de *Collins-Kimball*<sup>6,77</sup> e a formulação de *Kramer*<sup>78</sup>.

#### Formulação de Collins-Kimball

O tratamento de reações químicas na fase líquida deve contabilizar o efeito do solvente, considerando a capacidade dos reagentes de se difundirem e conduzi-los a colisões reativas efetivas. De acordo com o modelo de campo de reação de *Onsager*<sup>79-81</sup>, o solvente cria uma camada de solvatação, como uma gaiola em torno das entidades moleculares que participam da reação. O processo reativo entre as entidades moleculares A e B é representado pela equação sequencial:



onde o símbolo "solv" indica que a entidade molecular é cercada por uma camada de solvatação quase esférica com um raio específico.  $k_{DD}$  é a constante cinética de

difusão direta, na qual o reagente A se difunde pelo solvente até encontrar B para formarem os produtos;  $k_{DI}$  é a constante cinética de difusão da reação inversa; e  $k$  denota a constante cinética devido a colisões reativas efetivas: pode ser estimada a partir das Equações (1) ou (11). Assumindo a aproximação do estado estacionário<sup>81</sup> para a Equação (20), a expressão de *Smoluchowski* para a constante cinética de difusão<sup>82</sup>, assume a forma da Equação (21):

$$k_D = 4\pi r_{AB} D_{AB} \quad (21)$$

onde  $r_{AB}$  é a distância radial de reação e  $D_{AB}$  é a soma das constantes de difusão para cada reagente no solvente e  $k_D$  é a constante cinética de difusão. A generalização de *Collins* e *Kimball*<sup>76,77</sup> para reações bimoleculares irreversíveis controladas por difusão com constante cinética infinita,  $k_{DD} \cong k_{DI}$ <sup>83</sup>, gera uma constante de reação global observada,  $k_{obs}$ , expressa na Equação (22).

$$\frac{1}{k_{obs}} = \frac{1}{k_{TST}} + \frac{1}{k_D} \quad (22)$$

No código, a difusão é contabilizada pela formulação de *Stokes-Einstein*, Equação (23):

$$D_{AB} = \frac{k_B T}{6\pi\eta r_{AB}} \quad (23)$$

onde  $r_{AB}$  é o raio da camada de solvatação; e  $\eta$  a viscosidade do solvente. A viscosidade do solvente ( $\eta$ ) depende da temperatura e é estimada através da formulação de *Aquilanti-Mundim*, Equação (12), dando origem à Equação (24)<sup>84</sup>.

$$\eta(T) = \eta_0 \left( 1 + \frac{d\varepsilon^\ddagger}{RT} \right)^{1/d} \quad (24)$$

Aqui, o fator  $\eta_0$  é introduzido como uma contrapartida do fator pré exponencial  $A$  da formulação de *Arrhenius*, Equação (2). Mais informações sobre este formalismo no trabalho de *Carvalho-Silva* e colaboradores (2019)<sup>84</sup>.

### **Formulação de Kramer**

Para explicar os efeitos dinâmicos do solvente em um processo reativo e generalizar para processos unimoleculares e pseudo-unimoleculares, o modelo de *Kramer* considera um movimento estocástico do sistema, onde o efeito do solvente é adicionado considerando os movimentos brownianos ao longo do caminho da reação<sup>81</sup>.

Assumindo que a constante de atrito  $\mu$  é independente do tempo, a constante cinética da reação global  $k_{obs}$  pode ser calculada pela Equação (25):

$$k_{obs} = \kappa_{Kr} k_{TST} \quad (25)$$

na qual  $\kappa_{Kr}$  é o fator de transmissibilidade obtido por Kramer<sup>78</sup> expresso na Equação (26):

$$\kappa_{Kr} = \frac{1}{\omega^\ddagger} \left( \sqrt{\frac{\mu^2}{4} + \omega^\ddagger^2} - \frac{\mu}{2} \right) \quad (26)$$

para a constante de reação da teoria do estado de transição  $k_{TST}$  e suas variantes. Na Equação (26),  $\omega^\ddagger$  é a frequência imaginária do estado de transição e a constante de atrito é dada por  $\mu = (6\pi r_{AB}/M)\eta$ , onde  $r_{AB}$  e  $M$  são o raio da camada de solvatação e a massa molecular do estado de transição, respectivamente. Novamente, a viscosidade é calculada pela Equação (24).

### 1.3 Processo de Otimização via GSA

O processo de otimização é um algoritmo de ML no qual um valor ótimo de uma quantidade é determinado de acordo com uma série de parâmetros. Ou seja, seu objetivo é identificar os valores das variáveis de um processo (ou função) que irá fornecer o melhor resultado. Esse resultado é então avaliado por um critério de julgamento pré-estabelecido por uma função custo. Apesar desta definição simplista, os processos de otimização são de grande importância na comunidade científica e existem diversas dificuldades associadas a eles. Existem diversas variantes de algoritmos de otimização, tais como: método de *Newton-Raphson*, método do gradiente conjugado e o método estocástico do gradiente descendente.

Em termos matemáticos, um processo de otimização pode ser representado da seguinte maneira: dada uma função custo  $f: \vec{x} \rightarrow \mathbb{R}$  de algum conjunto de números reais  $\vec{x}$ , realiza-se uma minimização ou uma maximização:

- Minimização: procure um elemento do conjunto  $x_i \in \vec{x}$  tal que  $f(x_i) \leq f(x_j)$  para  $x_j \in \vec{x}$
- Maximização: procure um elemento do conjunto  $x_i \in \vec{x}$  tal que  $f(x_i) \geq f(x_j)$  para  $x_j \in \vec{x}$

Os processos de otimização utilizados em nossos códigos são baseados em um método de otimização estocástico chamado de “Resfriamento Simulado” (SA, do

inglês *Simulated Annealing*). O método SA tem sido amplamente usado devido à sua adequação para otimização em uma grande escala de problemas, especialmente para aqueles em que o mínimo global desejado está oculto entre muitos mínimos locais. O aspecto básico do método SA é que ele é análogo a termodinâmica, especialmente no que diz respeito à maneira com que os líquidos congelam e cristalizam ou com que os metais resfriam. A primeira solução não trivial a respeito deste método foi fornecida por *Kirkpatrick*<sup>85</sup> para sistemas clássicos e estendido por *Ceperley*<sup>86</sup> para sistemas quânticos. Esta solução segue estritamente a estatística de quase equilíbrio de *Boltzmann-Gibbs*, usando uma distribuição de visitação Gaussiana, e é algumas vezes referida como “Resfriamento Simulado Clássico” (CSA, do inglês “*Classical Simulated Annealing*”) ou máquina de *Boltzmann*. Outro estudo notável sobre o assunto foi a proposta de *Szu e Hartley*<sup>87</sup> de utilizar a distribuição de visitação de *Cauchy-Lorentz* ao invés da distribuição Gaussiana. Este algoritmo é conhecido como “Resfriamento Simulado Rápido” (FSA, do inglês “*Fast Simulated Annealing*”) ou máquina de *Cauchy*. Em 1996, *Tsallis* e colaboradores propuseram uma nova abordagem computacional chamada “Resfriamento Simulado Generalizado”, o GSA<sup>7</sup>, utilizada nos códigos produzidos neste trabalho.

O ponto central do GSA recai sobre a estatística de *Tsallis*, que, com base na teoria dos fractais múltiplos, propôs uma generalização da entropia. A entropia de *Tsallis* é representada pela Equação (27):

$$S_q = c \frac{1 - \sum_i p_i^q}{q - 1} \quad (27)$$

na qual  $q \in \mathbb{R}$ ,  $p_i^q$  é a probabilidade de configurações microscópicas e  $c$  é uma constante positiva. No limite de  $q \rightarrow 1$ , a equação da entropia de *Tsallis* é reduzida a equação da entropia de *Boltzmann-Gibbs*  $S = \lim_{q \rightarrow 1} S_q = c \sum_{p_i}^W p_i \ln p_i$ , na qual  $p_i$  é descrito na Equação (28).

$$p_i = \frac{[1 - \beta(1 - q)f_i]^{1/(1-q)}}{\sum_{i=1}^{\infty} [1 - \beta(1 - q)f_i]^{1/(1-q)}} \quad (28)$$

A variável representada por  $\beta$  é o multiplicador Lagrangiano igual a  $1/k\tau$ , e  $f_i$  é a função custo ou o espectro de energia do sistema. No modelo generalizado, a probabilidade da distribuição de visitação do GSA ( $g_q$ ) é expressa na Equação (29):

$$g_{q_V}(\Delta x_t) = \left(\frac{q_V-1}{\pi}\right)^{D/2} \frac{\Gamma\left[\frac{1}{q_V-1} + \frac{D-1}{2}\right]}{\Gamma\left[\frac{1}{q_V-1} - \frac{1}{2}\right]} \frac{[\tau_{q_T}(t)]^{-\frac{D}{3-q_V}}}{\left\{1 - (q_V-1) \frac{(\Delta x_t)^2}{[\tau_{q_T}(t)]^{\frac{-2}{3-q_V}}}\right\}^{\frac{1}{q_V-1} + \frac{D-1}{2}}} \quad (29)$$

na qual  $D$  é o número de dimensões ou componentes de  $x_t$ ,  $q$  é um parâmetro ajustável,  $\tau_{q_T}(t)$  é a temperatura de visitação no tempo  $t$ , que é um número discreto que representa o passo da iteração computacional e  $\Gamma$  é a função *gamma*. Nesta metodologia,  $\tau_{q_T}(t)$  é uma temperatura artificial que é gradualmente resfriada, fazendo analogia à técnica de resfriamento frequentemente utilizado na metalurgia quando um metal fundido atinge seu estado cristalino (mínimo global da energia termodinâmica). No nosso caso, a temperatura é considerada um ruído externo. O “resfriamento” da temperatura artificial  $\tau_{q_T}(t)$  é expresso na Equação (30).

$$\tau_{q_T}(t) = \tau_{q_T}(1) \frac{2^{q_T-1} - 1}{(1+t)^{q_T-1} - 1} \quad (30)$$

A probabilidade de aceitação é dada na Equação (31).

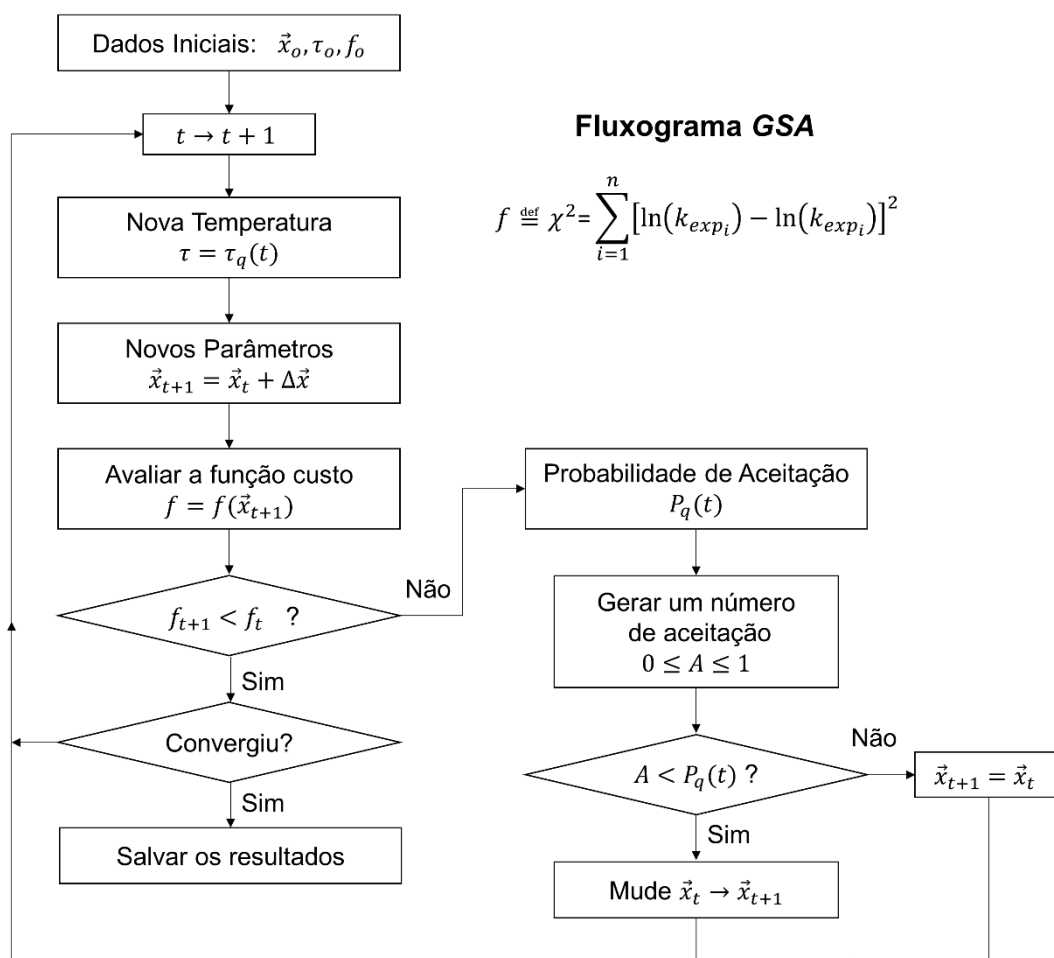
$$P_{q_A}(x_t \rightarrow x_{t+1}) = \frac{1}{1 + \left\{1 + (q_A - 1) \frac{f_{t+1} - f_t}{\tau_{q_T}(t)}\right\}^{\frac{1}{q_A-1}}} \quad (31)$$

Em suma, o método GSA (Figura 2) é um método de ML baseado na dinâmica estocástica que permite a identificação de um mínimo global na hiper superfície de energia. Em uma publicação de De Andrade e colaboradores<sup>88</sup>, os autores apresentaram provas de que o algoritmo GSA sempre convergirá ao mínimo absoluto do sistema, com três parâmetros independentes:  $q_A$ ,  $q_V$  e  $q_T$ , representando respectivamente as funções de probabilidade de aceitação, distribuição de visitação e temperatura.

É importante pontuar que sempre que citada neste trabalho a dinâmica estocástica de ajuste do GSA, o processo envolvido é baseado na minimização global do método estatístico padrão  $\chi^2$ , que é definido na Equação (32):

$$\chi^2 = \sum_{i=1}^n \left[ \ln(k_{exp_i}) - \ln(k_{teo_i}) \right]^2 \quad (32)$$

na qual  $n$  é o número de dados experimentais,  $k_{exp_i}$  é o  $i$ -ésimo valor da constante cinética experimental e  $k_{teo_i}$  é o  $i$ -ésimo valor da constante cinética teórica (ajustada), como descrito nas Equações (1) e (11) dos modelos de *Arrhenius* e *d-Arrhenius*.



**Figura 2:** Fluxograma básico para construção do algoritmo GSA. Como discutido previamente, o GSA é um método de ML estocástico que resolve problemas de otimização e é uma técnica probabilística para encontrar o mínimo global ótimo de uma dada função custo.

## 1.4 Energia de Ativação e o Gráfico da Transitividade

As curvaturas no gráfico de *Arrhenius* representam uma dependência da temperatura na energia de ativação aparente  $E_a$ , cuja definição recomendada pela União Internacional de Química Pura e Aplicada (IUPAC, do inglês *International Union of Pure and Applied Chemistry*)<sup>41,89</sup> é expressa na Equação (33):

$$E_a = -R \frac{d \ln k(T)}{d(1/T)} = RT^2 \frac{d \ln k(T)}{d(T)} \quad (33)$$

na qual:  $R$  é a constante dos gases ideais,  $T$  é a temperatura absoluta; e  $k$  é a taxa da reação. Aplicando a formulação de *Arrhenius*, Equação (2), na Equação (33), verifica-se que a energia de ativação é constante e não depende da temperatura. Esta proposta é válida para processos físico-químicos, pelo menos no intervalo de temperatura de interesse. Entretanto, como abordado na seção 1.2.2, em diversos processos descritos na literatura existem desvios de linearidade no gráfico de



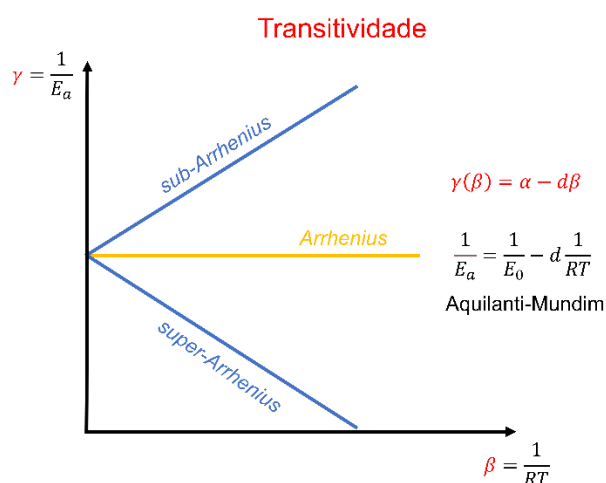
*Arrhenius*, nos quais a energia de ativação é dependente da temperatura. Para verificar esta dependência aplicamos a formulação de Aquilanti-Mundim, Equação (12), na Equação (33). A energia de ativação é então descrita pela Equação (34).

$$E_a = E_0 \left(1 - \frac{dE_0}{RT}\right)^{-1} \quad (34)$$

Considerando  $\alpha = 1/E_0$  e  $\beta = 1/RT$ , aqui nós introduzimos o inverso da energia de ativação aparente, a fim de definir a função de transitividade na Equação (35):

$$\gamma(\beta) \equiv \frac{1}{E_a(\beta)} = \alpha - d\beta \quad (35)$$

para construir um *plot* com escala apropriada,  $\gamma(\beta)$  vs.  $\beta$ , no qual curvaturas regulares no plano de *Arrhenius* são aproximadamente lineares: o “*Transitivity Plot*” (Figura 3). A função  $\gamma(\beta)$  pode ser interpretada como uma medida da tendência da ocorrência de uma reação. Esta função permite uma análise uniforme de processos experimentais e teóricos que envolvem a constante cinética química, como tunelamento químico, propriedades de transporte e difusão nas vizinhanças de transição de fases. O aparato teórico para conectar a função de transitividade,  $\gamma(T)$ , e as formulações fenomenológicas da constante cinética e vice-versa é amplamente discutido na publicação de V. H. Carvalho-Silva e colaboradores (2019)<sup>84</sup> e nas referências nela contidas.



**Figura 3:** Representação genérica do gráfico da transitividade  $\gamma(\beta) = 1/E_a$  vs.  $\beta = 1/RT$ . O comportamento *Arrhenius* é dado pela linha amarela paralela ao eixo  $\beta$  e os desvios sub e super *Arrhenius* são representados nas linhas azuis.

A introdução do gráfico da transitividade aqui proposto, definido como o plano  $\gamma(\beta)$  vs.  $\beta$ , serve para dar um significado geométrico aos parâmetros

fenomenológicos que ocorrem no estudo de comportamentos não-*Arrhenius*, como mostrado na Figura 3.

## 1.5 Exemplos e aplicações

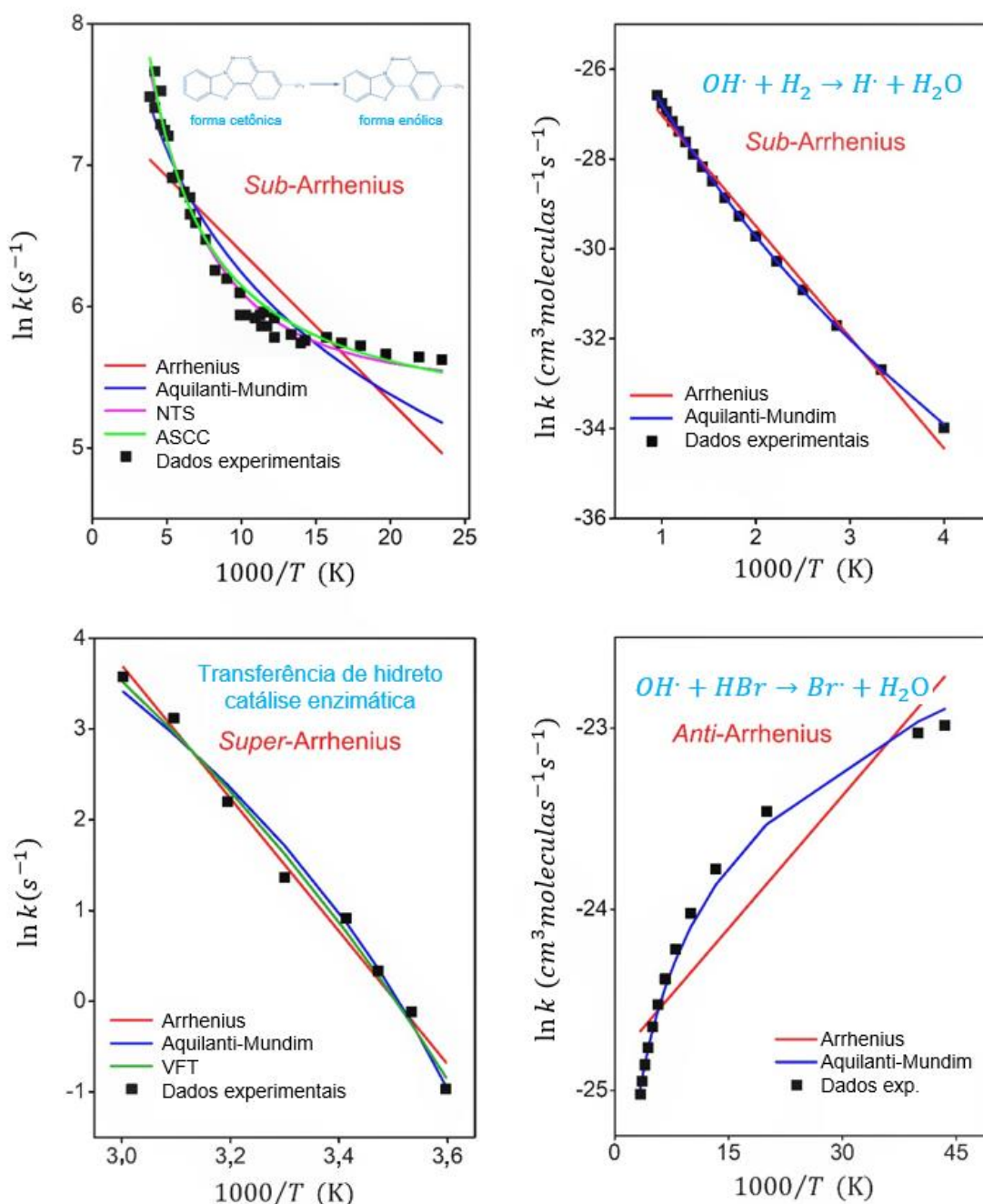
### 1.5.1 Ajustes de função – Gráficos de *Arrhenius*

Para ilustrar o uso das funcionalidades e validar a precisão e o funcionamento do nosso código, foram realizados ajustes de constantes cinéticas para quatro sistemas em regimes diferentes: *sub-Arrhenius* com tunelamento profundo, *sub-Arrhenius* com tunelamento moderado, *super-Arrhenius* e *anti-Arrhenius*. Para os casos estudados foram realizados diferentes tipos de ajustes disponíveis no código utilizando todas as formulações cinéticas disponíveis. Os melhores ajustes foram então definidos como aqueles que apresentaram menores valores para a função custo  $\chi^2$ , Equação (32), e somente esses resultados serão aqui apresentados juntamente com o ajuste convencional de *Arrhenius*, tanto na Figura 4 quanto na Tabela 1.

O primeiro exemplo diz respeito a tautomerização ceto-enólica do 2-(2'-hidroxi-4'-metilfenil) benzoxazol (MeBO), um caso bem conhecido de regime de tunelamento profundo<sup>35</sup>. No segundo exemplo, a reação  $OH\cdot + H_2 \rightarrow H\cdot + H_2O$ <sup>90</sup> foi avaliada para analisar o comportamento *sub-Arrhenius* sob regime de tunelamento moderado. O terceiro exemplo é centrado nas investigações que revelaram o comportamento *super-Arrhenius* para constantes cinéticas de processos promovidos por catálise enzimática<sup>91-93</sup>. Neste caso consideramos a constante cinética da reação catalisada de transferência de hidreto entre o substrato e  $NAD^+$ , o qual exibe uma forte curvatura no intervalo de temperatura entre 5 e 65 °C<sup>92</sup>. O quarto e último trata do comportamento *anti-Arrhenius*, o qual é caracterizado pela diminuição da constante cinética da reação com o aumento da temperatura: a reação  $OH\cdot + HBr \rightarrow Br\cdot + H_2O$  é um caso típico, tanto do ponto de vista teórico quanto experimental exibe dependência negativa da temperatura com a constante cinética da reação<sup>94-96</sup>.

A Figura 4 mostra os dados experimentais e os gráficos de ajuste no gráfico de *Arrhenius* para os 4 sistemas citados. Os parâmetros de ajuste e a análise estatística da qualidade dos ajustes para todos os sistemas são fornecidos na Tabela 1. É importante ressaltar que o  $\chi^2$  é uma medida estatística que compara a

qualidade do ajuste de acordo com os valores de referência, e, em um processo de otimização afim de obter parâmetros que tenham significado físico, é necessário evitar efeitos de compensação<sup>97</sup>, pois o ajuste das formulações pode levar a múltiplas soluções. Para isso, o usuário deve explorar o seu conhecimento prévio do sistema e escolher melhores valores iniciais ao ajuste, além de selecionar os melhores parâmetros do otimizador GSA para seu caso.



**Figura 4:** Gráficos de Arrhenius comparando as constantes cinéticas experimentais e ajustadas para: reação de tautomerização ceto-enólica (comportamento sub-Arrhenius sob tunelamento profundo), reação  $\text{OH}\cdot + \text{H}_2 \rightarrow \text{H}_2\text{O} + \text{H}\cdot$  (comportamento sub-Arrhenius sob regime de tunelamento moderado), transferência de hidreto com catálise enzimática (comportamento super-Arrhenius) e reação  $\text{OH}\cdot + \text{HBr} \rightarrow \text{Br}\cdot + \text{H}_2\text{O}$  (comportamento anti-Arrhenius). As referências dos dados experimentais serão apresentadas na Tabela 1.

No primeiro caso, referente a reação de tautomerização ceto-enólica de comportamento *sub-Arrhenius* sob regime de tunelamento profundo, as formulações NTS e ASCC foram satisfatórias para realizar o ajuste enquanto Aquilanti-Mundim reproduziu o comportamento dos dados apenas em um intervalo específico de temperatura.

No segundo caso, sob um regime *sub-Arrhenius* de tunelamento moderado, a formulação AM foi uma excelente opção para descrever a reação  $OH\cdot + H_2$ .

No terceiro caso, para reproduzir o comportamento *super-Arrhenius*, as formulações AM e VFT se mostraram as melhores opções.

Para o último caso, os parâmetros da formulação de AM obtidos para a reação  $OH\cdot + HBr$  indicam que esta é a melhor opção para reproduzir o comportamento *anti-Arrhenius*.

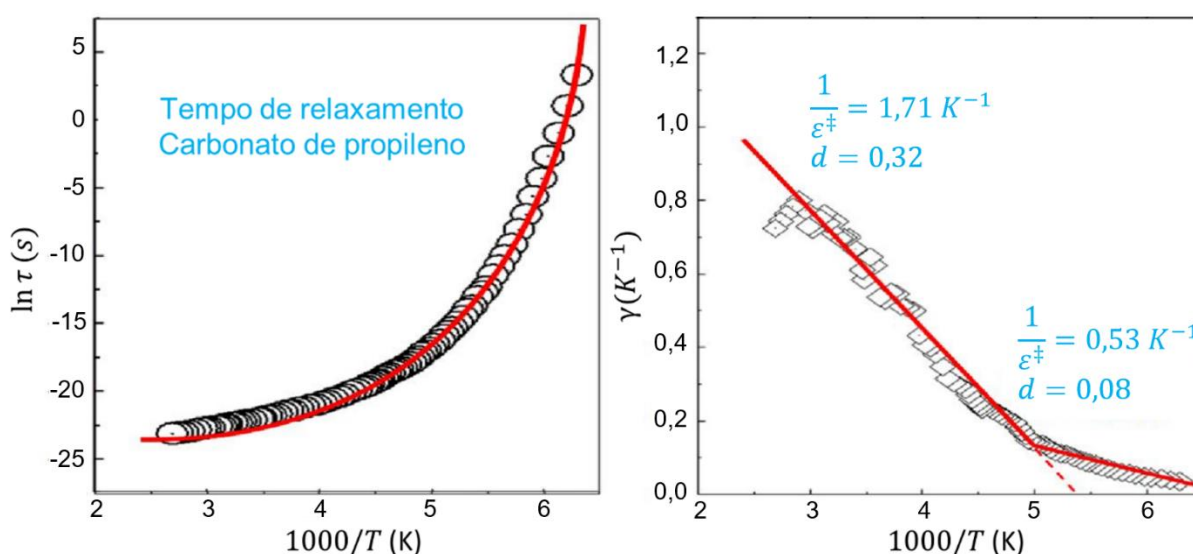
Como esperado, a formulação de *Arrhenius* é claramente inadequada para explicar os desvios a baixa temperatura para todas as reações apresentadas.

**Tabela 1:** Parâmetros ajustados para as formulações AM, ASCC, NTS e VFT utilizando o código *Transitivity* para os processos de Tautomerização ceto-enólica<sup>35</sup>, reação  $OH \cdot + H_2$ <sup>90</sup>, catálise enzimática<sup>92</sup> e reação  $OH \cdot + Br$ <sup>98</sup>. Energy ( $E_a$ ,  $\varepsilon^\ddagger$ ,  $E_v$  e  $E_0$ ) estão em *cal/mol* e a temperatura ( $T_0$  e  $B$ ) em Kelvin (*K*). As unidades dos fatores pré-exponenciais podem ser obtidas em suas referências.

Formulações	Parâmetros Ajustados	Processos Químicos			
		Tautomerização Ceto-Enólica <sup>35</sup> <i>sub</i> -Arrhenius (tunelamento profundo)	$OH \cdot + H_2 \rightarrow H \cdot + H_2$ <sup>90</sup> <i>sub</i> -Arrhenius (tunelamento moderado)	Catálise enzimática <sup>92</sup> Super-Arrhenius	$OH \cdot + HBr \rightarrow Br \cdot + H_2O$ <sup>98</sup> Anti-Arrhenius
Arrhenius $k(T) = Ae^{\left(\frac{-E_a}{k_B T}\right)}$	$A$	$1,74 \times 10^3$	$2,16 \times 10^{-11}$	$1,52 \times 10^{11}$	$1,66 \times 10^{-11}$
	$E_a$	214	4891	14600	-94,6
	$\chi^2$	$1,10 \times 10^{-2}$	$4,20 \times 10^{-3}$	$2,60 \times 10^{-2}$	$6,69 \times 10^{-2}$
Aquilanti-Mundim (AM) $k(T) = A \left(1 - d \frac{\varepsilon^\ddagger}{k_B T}\right)^{\frac{1}{d}}$	$A$	$3,32 \times 10^6$	$1,11 \times 10^{-10}$	$1,91 \times 10^4$	$7,43 \times 10^{-14}$
	$\varepsilon^\ddagger$	318,06	9170	2391	-324,61
	$d$	-0,81	-0,086	0,207	1,24
Aquilanti-Sanchez-Coutinho-Carvalho (ASCC) $k(T) = A \left(1 - d \frac{\varepsilon^\ddagger}{k_B T + E_v}\right)^{\frac{1}{d}}, d = \frac{-1}{3} \left(\frac{E_v}{2\varepsilon^\ddagger}\right)^2$	$A$	$2,33 \times 10^4$	-	-	-
	$\varepsilon^\ddagger$	2441	-	-	-
	$E_v$	429	-	-	-
Sato-Nakamura-Takayanagi(NTS) $k(T) = Ae^{\left(\frac{-E_0}{k_B(T^2+T_0^2)^{\frac{1}{2}}}\right)}$	$A$	$3,12 \times 10^4$	-	-	-
	$E_0$	1655	-	-	-
	$T_0$	168	-	-	-
Vogel-Fulcher-Tamman (VFT) $k(T) = Ae^{\left(\frac{B}{(T-T_0)}\right)}$	$\chi^2$	$7,38 \times 10^{-3}$	-	-	-
	$A$	-	-	$1,25 \times 10^{-5}$	-
	$B$	-	-	-1298	-
	$T_0$	-	-	175	-
	$\chi^2$	-	-	$2,16 \times 10^{-2}$	-

### 1.5.2 Ajustes de função – Gráfico da transitividade

As curvaturas em processos não-*Arrhenius* observadas na seção anterior (1.5.1) são facilmente identificadas através dos gráficos de *Arrhenius*, entretanto, ajustar o comportamento de formulações fenomenológicas pode levar a múltiplas soluções que dificultam a interpretação física dos parâmetros obtidos, como discutido previamente na mesma seção. A função da transitividade aqui proposta tem o intuito de linearizar estas curvaturas observadas no gráfico de *Arrhenius* de forma análoga a função da transitividade proposta na década de 1980 para avaliar a transição vítrea em materiais super-resfriados<sup>99–101</sup>. Para ilustrar esta proposta foi avaliado o processo de relaxamento do carbonato de propileno<sup>100</sup>. A Figura 5 mostra o gráfico de *Arrhenius* (à esquerda) e o gráfico da Transitividade (à direita) para a dependência da temperatura com o tempo de relaxação do carbonato de propileno.



**Figura 5:** Gráficos de *Arrhenius* (superior) e da transitividade (inferior) para a dependência da temperatura do tempo de relaxamento do carbonato de propileno. Os símbolos de diamante representam os valores de transitividade obtidos numericamente e suavizados com o filtro *Savitzky – Golay*. As linhas vermelhas enfatizam duas regiões onde a dependência da transitividade com a temperatura é linearizada, conforme esperado pela lei de *Aquilanti-Mundim*.

Na Figura 5, referente ao relaxamento do carbonato de propileno, a curvatura no gráfico de *Arrhenius* assume um comportamento linear quando observada sob a perspectiva do gráfico da transitividade. No entanto, a 198K o comportamento linear sofre uma mudança na sua inclinação e uma transição para outro regime é observada – uma quebra não percebida no gráfico de *Arrhenius*.

Com a linearização dos dados no gráfico da transitividade o ajuste se torna muito mais simples. Na primeira faixa de alta temperatura os parâmetros Aquilanti-Mundim são  $1/\varepsilon^\ddagger = 0,53K^{-1}$  e  $d = 0,08$ , enquanto na segunda faixa para temperaturas abaixo de  $198K$  os parâmetros Aquilanti-Mundim são  $1/\varepsilon^\ddagger = 1,71K^{-1}$  e  $d = 0,32$ .

### 1.5.3 Estimativa de constantes cinéticas de reação

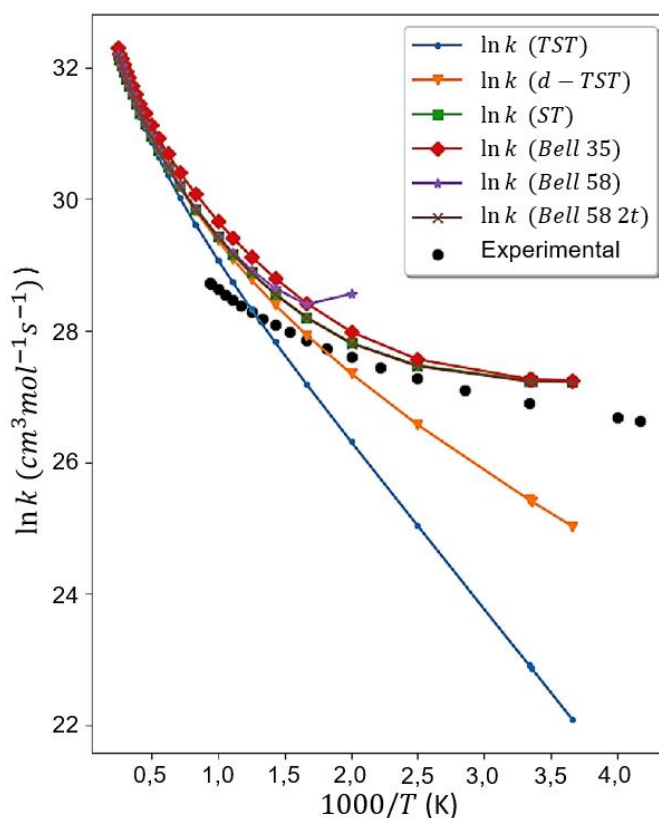
A estimativa de constante cinética da reação  $OH\cdot + HCl \rightarrow H_2O + Cl\cdot$  em fase gasosa foi realizada para validar a TST e as correções de tunelamento implementadas no código *Transitivity*. Além disso, usando a sugestão do código *Eyringpy*<sup>57</sup>, a reação  $NH_3 + OH\cdot \rightarrow NH_2\cdot + H_2O$  foi escolhida para demonstrar a precisão dos modelos de *Collins-Kimball* e *Kramer* para estimar as constantes cinéticas das reações em fase aquosa.

#### **Reação $OH\cdot + HCl \rightarrow H_2O + Cl\cdot$**

A constante cinética da reação entre o radical hidroxila e o cloreto de hidrogênio é apenas ligeiramente dependente da temperatura na faixa de  $138 - 300K$ , embora à medida que a faixa aumenta ainda mais, uma dependência da temperatura é observada. Esta forte curvatura côncava detectada no gráfico de *Arrhenius* (comportamento sub-*Arrhenius*) é uma evidência muito convincente do papel de tunelamento quântico profundo para esta reação<sup>68</sup>. Foram empregadas as correções de tunelamento *d*-TST, Bell35, Bell58 e ST para calcular as constantes de taxa para a reação de  $OH\cdot + HCl$  em uma ampla faixa de temperaturas ( $200 - 2000K$ ). As propriedades de estrutura eletrônica dos reagentes, dos produtos e do estado de transição foram calculadas empregando-se o nível de cálculo MP2/aug-cc-pVDZ usando o programa Gaussian09<sup>102</sup>. Um estudo completo desta reação usando nossa metodologia pode ser encontrado na publicação de N. D. Coutinho e colaboradores (2018)<sup>103</sup>.

A Figura 6 foi obtida através do programa *Transitivity* e mostra uma comparação entre as diversas constantes cinéticas calculadas e os dados experimentais obtidos por Ravishankara e colaboradores<sup>104</sup>. Nenhuma diferença significativa foi encontrada usando as correções de tunelamento de Bell58 e ST. Por outro lado, como esperado, uma divergência é observada na formulação de Bell58

em  $T_c$  e a  $d$ -TST falha em descrever o comportamento na região de baixas temperaturas, onde o regime de tunelamento profundo é dominante. Isto mostra que sua validade é limitada a situações nas quais o tunelamento é fraco. A TST tradicional é apresentada de modo comparativo e demonstra a necessidade de correções.



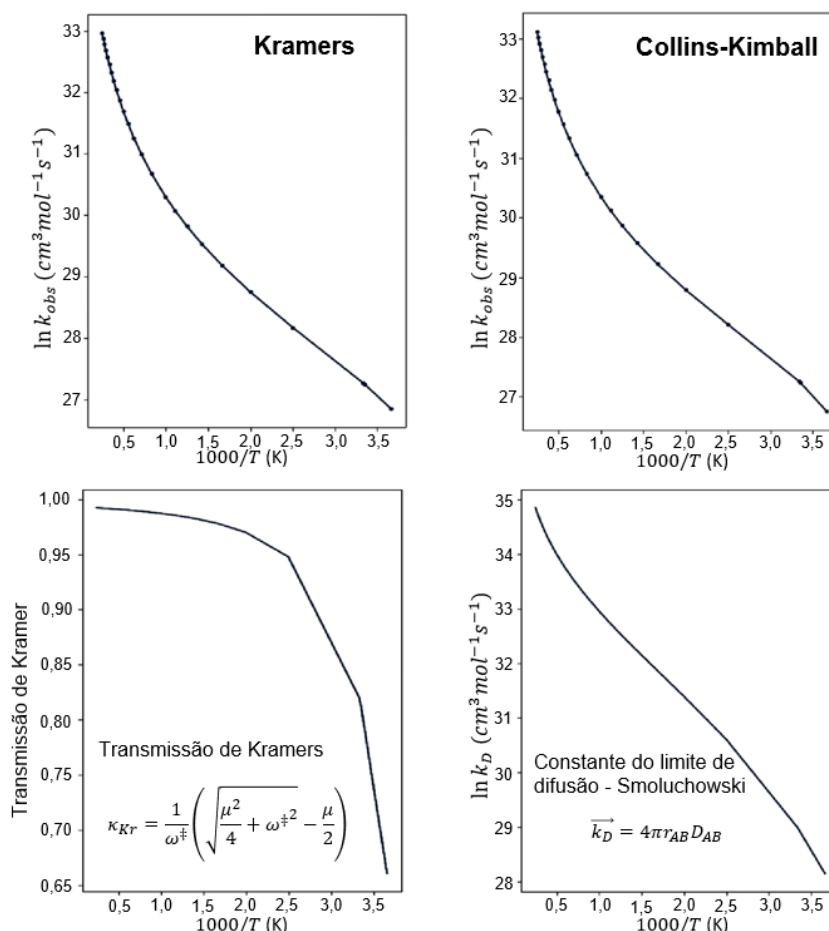
**Figura 6:** Gráfico de Arrhenius obtido a partir do código Transitivity para a reação  $OH\cdot + HCl \rightarrow Cl\cdot + H_2O$  utilizando a TST com as correções de Bell35, Bell58, ST e para  $d$ -TST. Os dados experimentais de Ravishankara e colaboradores<sup>104</sup> estão disponíveis para comparação e foram exibidos como pontos inteiros

### Reação $NH_3 + OH\cdot \rightarrow NH_2\cdot + H_2O$

A reação  $NH_3 + OH\cdot \rightarrow NH_2\cdot + H_2O$  nos permite ilustrar a acurácia da metodologia na fase líquida. Energias, geometrias e frequências do ponto estacionário foram extraídas no mesmo nível de cálculo usado no programa *Eyringpy*<sup>57</sup>. Os gráficos na parte superior da Figura 7 mostram a dependência da temperatura com a constante cinética da reação global,  $k_{obs}$ , para a reação  $NH_3 + OH\cdot \rightarrow NH_2\cdot + H_2O$  no intervalo de 273 a 4000K de temperatura utilizando os modelos de *Kramer* e *Collins-Kimball*. A constante de difusão de *Smoluchowski*  $k_D$



que avalia o limite de difusão para processos reativos bimoleculares, incluindo o efeito de solvente, é apresentada nos gráficos na parte inferior da Figura 7.



**Figura 7:** Na parte superior estão os gráficos de Arrhenius obtidos pelo código *Transitivity* para a reação  $NH_3 + OH^\cdot \rightarrow NH_2^\cdot + H_2O$  utilizando as formulações de *Kramer* e *Collins-Kimball*. Na parte inferior estão os gráficos do coeficiente de transmissão de Kramer e da constante do limite de difusão de *Smoluchowski* como uma função inversa da temperatura.

O coeficiente de transmissão de Kramer, que avalia a interferência do efeito de atrito do solvente no processo reativo como uma função da temperatura, é mostrado no gráfico inferior esquerdo da Figura 7. A uma temperatura de 298,15K a formulação de *Kramer* apresentou um valor de  $6,73 \times 10^{11} cm^3 mol^{-1} s^{-1}$  para a constante cinética da reação, enquanto o valor obtido pela formulação de *Collins-Kimball* foi de  $6,77 \times 10^{11} cm^3 mol^{-1} s^{-1}$  (o valor experimental é  $\sim 10^{11} cm^3 mol^{-1} s^{-1}$ <sup>105-107</sup>). O valor de  $k_D$  por *Smoluchowski* (*Collins-Kimball*) é  $3,73 \times 10^{12} cm^3 mol^{-1} s^{-1}$ , em acordo com o calculado por *Dzib* e colaboradores<sup>57</sup>:  $\vec{k}_D = 3,60 \times 10^{12} cm^3 mol^{-1} s^{-1}$ .

## 1.6 Síntese de chalconas – uma aplicação do *Transitivity*

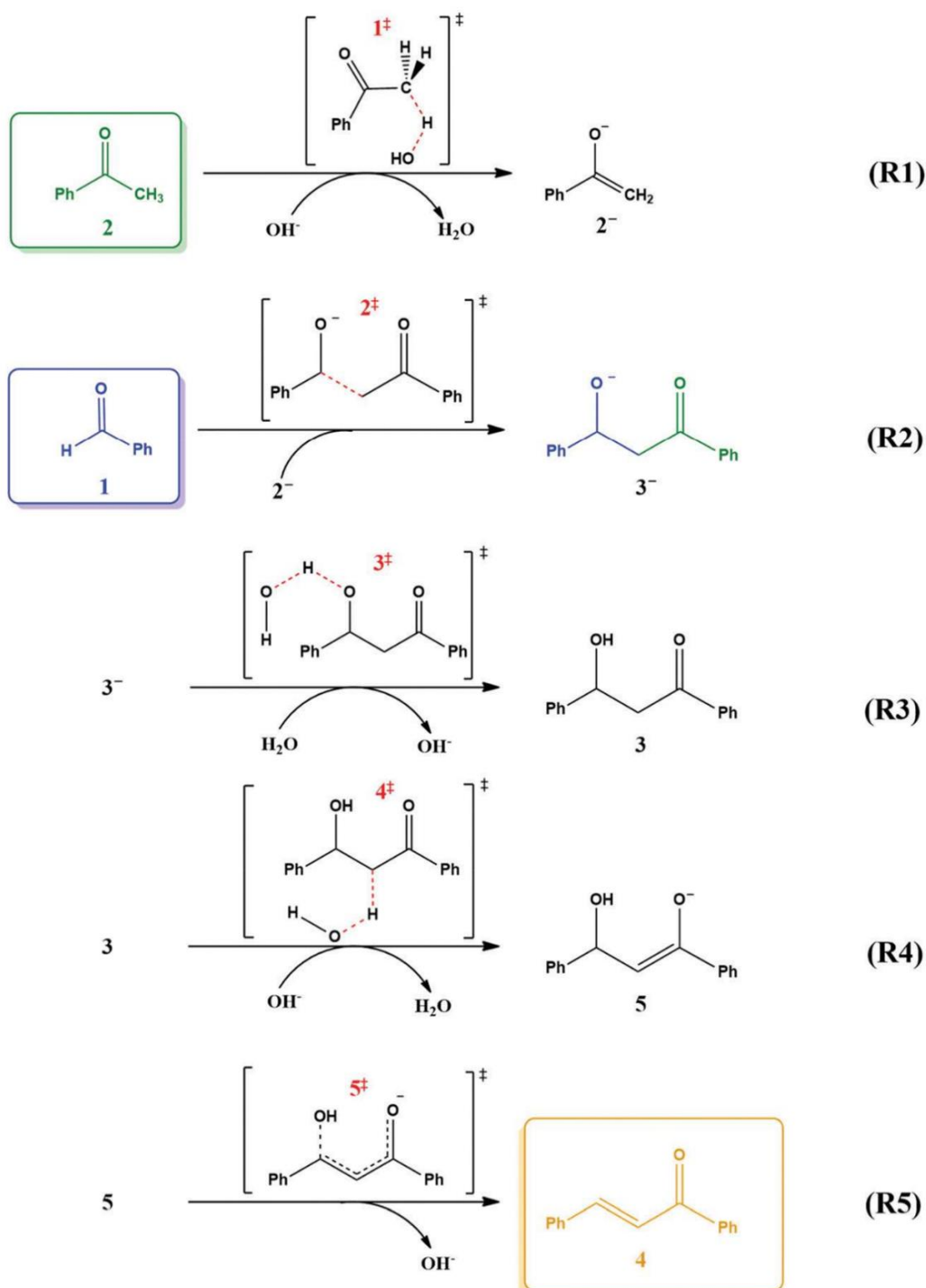
Na seção 1.5 foram apresentadas uma gama de exemplos e aplicações possíveis ao programa *Transitivity*. As situações apresentadas e os dados utilizados foram retirados da literatura a fim de ilustrar detalhadamente cada função do programa. Nesta seção será apresentada uma análise completa do diagrama do perfil de energia livre de *Gibbs* na síntese de chalconas pelo mecanismo de *Claisen-Schmidt*<sup>108</sup>. Todos os cálculos e análises foram realizados com o código *Transitivity*.

A reação de *Claisen-Schmidt* é um protocolo bem conhecido para a síntese de chalconas<sup>109,110</sup> e pode ser realizada de forma eficiente pela condensação catalisada por base do benzaldeído (**1**) com acetofenona (**2**), um caso representativo clássico da condensação aldólica<sup>111,112</sup> para a formação de chalconas (**4**)<sup>113–115</sup>. A Figura 8 mostra as etapas-chave amplamente aceitas para o mecanismo de condensação de *Claisen-Schmidt*.

As várias etapas podem ser limitadas por aspectos cinéticos e termodinâmicos controlando a constante de velocidade global na condensação aldólica. Estudos objetivando elucidar as principais etapas deste mecanismo vem sendo realizados desde o início do século passado, em 1940 *Coombs* e *Evans*<sup>116</sup> provaram a lei cinética para o mecanismo e consolidaram durante décadas a compreensão de que a formação da ligação C-C, a etapa **R2**, é a Etapa Lenta de Controle (RSC, do inglês: *Rate Controlling Step*) da constante cinética. Entretanto, em 2016 *Chang* e *Perrin* estabeleceram o mecanismo completo para a condensação aldólica<sup>117</sup> e provaram experimentalmente que a etapa **R5**, formação da ligação C=C, é a RSC, levando a uma revisão do entendimento que se tinha até então. Para mais detalhes acerca das etapas do mecanismo reacional ver Figura 8.

*Chang* e *Perrin* observaram um forte efeito isotópico do solvente ao analisar, em meio aquoso, a propensão do cetol **3** em se reverter aos reagentes **1** e **2** ou a progredir ao produto **4**. No entanto, os autores indicaram que algumas restrições foram necessárias à dinâmica experimental, como uma adição de acetonitrila ( $CH_3CN$ ) com cossolvente na síntese, atribuindo uma característica mista ao solvente e podendo sobrepor efeitos isotópicos e de solvente. Assim como no trabalho de *Chang* e *Perrin*, os diversos estudos experimentais já realizados acerca deste mecanismo geralmente apresentam restrições experimentais associadas a utilização de solventes e catalisadores específicos<sup>118–121</sup>. Estas restrições

experimentais são impostas devido a complexidades do processo de síntese e provavelmente são as responsáveis por tantos resultados contraditórios ao longo do tempo. Por isso, estudos teóricos sobre o tema se fazem necessários.



**Figura 8:** Reagentes, intermediários e estruturas de transição e todas as etapas elementares da formação de chalcona 4 catalisada por base a partir de benzaldeído 1 e acetofenona 2: (R1) primeira enolização, (R2) formação da ligação C-C, (R3) equilíbrio prótico, (R4) segunda enolização, e (R5) eliminação de hidróxido e formação da ligação C=C. A nomenclatura e indexação utilizada no trabalho de Chang e Perrin<sup>117</sup> foram mantidas para facilitar comparações.

Diante disso, estudos teóricos vêm sendo realizados recentemente a fim de esclarecer os detalhes acerca deste mecanismo<sup>122,123</sup>. Aqui, equipado com um aparato químico computacional, os efeitos de solvente e do isótopo ponderados pela contribuição da temperatura e do atrito do solvente são individualizados, ampliando o papel decisivo desses fatores físico-químicos na topografia do diagrama de energia livre. Serão realizadas considerações adicionais acerca do tunelamento quântico na reatividade da condensação de *Claisen-Schmidt*, uma vez que um número crescente de estudos<sup>124–126</sup> enfatizou a importância de incluir esta correção em reações químicas orgânicas, principalmente quando há etapas de transferência de hidrogênio.

### 1.6.1 Procedimentos Computacionais

Os pontos estacionários das etapas elementares detalhadas na Figura 8 foram otimizados empregando o nível de cálculo do modelo composto B2PLYPD/AUG-cc-pVDZ // B3LYP/6-31+G(3df,3dp)<sup>127,128</sup>. As correções de solvatação SMD<sup>129</sup> e de dispersão de *Grimme*<sup>130</sup> foram aplicadas tanto na otimização no nível B3LYP quanto nos cálculos de energia *single point* no nível B2PLYP. Os cálculos teóricos de geometria e estrutura eletrônica foram realizados utilizando o programa *Gaussian16*<sup>102</sup>. As constantes cinéticas para as etapas de **R1** a **R5** (ver Figura 8) ao longo do intervalo de temperatura experimental (273 – 373 K) foram calculados pela teoria do estado de transição deformado<sup>61</sup> incluindo a correção de tunelamento quântico de *Skodje-Truhlar*<sup>75</sup>. A aproximação de *Smoluchowski*<sup>77</sup> foi aplicada na enolização controlada por difusão rápida<sup>113</sup>, etapa **R3**. Para contabilizar os efeitos dinâmicos do solvente nas etapas elementares a correção de transmissão de *Kramer*<sup>78</sup> foi adicionada às constantes cinéticas da reação, permitindo uma descrição mais realista do modelo de solvatação. A constante cinética bimolecular de transferência de prótons para as etapas inversas **R1**, **R3** e **R4** com uma molécula de água foram simplificadas em uma constante cinética unimolecular aplicando o mecanismo de *Eigen*<sup>83</sup>. Todas as análises e cálculos cinéticos foram realizados utilizando o código *Transitivity*<sup>31</sup>.

### 1.6.2 Influência do solvente no perfil da energia livre de *Gibbs*

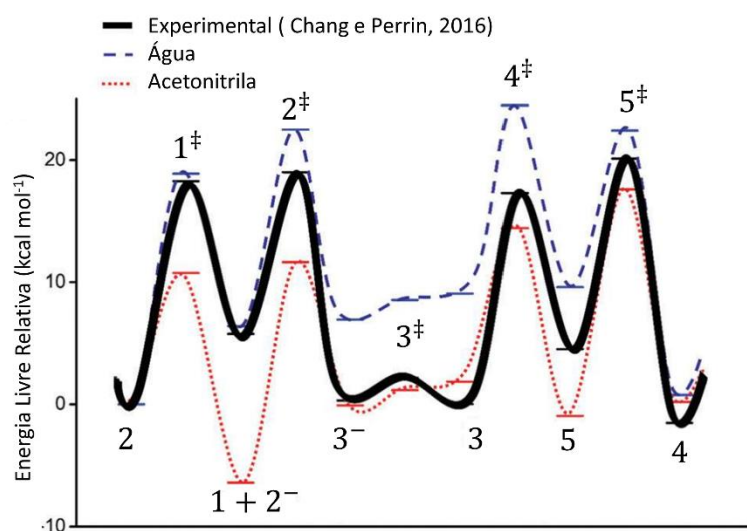
O perfil de energia livre da condensação de *Claisen-Schmidt* catalisada por base é apresentado na Figura 9, detalhando os pontos estacionários da Figura 8. A curva preta sólida mostra o perfil estimado experimentalmente por Chang e Perrin<sup>117</sup>, onde a etapa **R5** foi dada como a etapa lenta de controle cinético da reação. O perfil teórico da energia livre utilizando água ( $H_2O$ ) com solvente é representado pela linha azul tracejada na Figura 9, e os dados de energia para água pesada ( $D_2O$ ) são mostrados na Tabela 2. Embora a maioria dos estudos sobre a formação de chalconas<sup>116,119,131–133</sup> tenha sido realizada em solução aquosa de etanol, aqui não há consideração do etanol a ser simulado como solvente para evitar mal-entendidos devido a reações colaterais e para simplificar o sistema.

O estudo experimental mais representativo para a condensação de *Claisen-Schmidt* catalisada por base<sup>117,134</sup> em água exigiu a adição de quantidades consideráveis de acetonitrila ( $CH_3CN$ ) como um cossolvente a fim de manter a solubilidade dos substratos e do produto chalcona e permitir a reprodutibilidade do comportamento cinético. Portanto, na Figura 9 a linha pontilhada em vermelho mostra o perfil de energia livre considerando a acetonitrila ( $CH_3CN$ ) como solvente, obtido no mesmo nível de cálculo aplicado em  $H_2O$ .

A Figura 9 mostra uma mudança significativa na topografia de energia livre da formação de chalcona quando solventes polares próticos ( $H_2O$ , e  $\epsilon = 35,69$ ) são utilizados. Uma mudança repentina causada pela ação da acetonitrila como solvente é a conversão das etapas **R1** e **R4** em processos espontâneos ( $\Delta G < 0$ ), evitando a reversibilidade nessas etapas. Esse comportamento nas reações aldólicas já havia sido discutido anteriormente por *Nielsen* e *Houlihan*<sup>111</sup>, que estabeleceram a capacidade dos solventes apróticos em promover etapas irreversíveis, induzindo um estágio de equilíbrio excessivamente lento até a formação do produto<sup>135</sup>.

Outra manifestação substancial do efeito do solvente pode ser observada na estabilização das estruturas do estado de transição devido a acetonitrila. A tendência na estabilização das espécies **1**<sup>‡</sup> e **4**<sup>‡</sup> surge devido a fraca solvatação que a acetonitrila promove no ânion hidroxila, estimulando o processo de abstração de hidrogênio. Além disso, a propensão de solventes próticos, como água, para desestabilizar estruturas de ressonância enólicas<sup>136,137</sup> também deve ser considerada nas etapas **R1** e **R4** (ver Figura 9 e Tabela 2), permitindo um caminho

competitivo mediado pelo mecanismo de *Grotthuss* (salto prótico)<sup>123</sup>. Um efeito de solvente significativo nas barreiras intrínsecas das etapas **R1** e **R5** também pode ser observado, com o meio contínuo de acetonitrila auxiliando uma diminuição de  $8,16 \text{ kcal mol}^{-1}$  na primeira enolização e um aumento de  $5,73 \text{ kcal mol}^{-1}$  para a eliminação do ânion hidroxila e formação da ligação C=C (ver Tabela 2). A etapa **R4** sofreu uma diminuição da barreira intrínseca no meio contínuo acetonitrila, tendência já avaliada por *Bunnett*<sup>138</sup>, que atestou que um solvente mais apolar deve mover o mecanismo no sentido da eliminação do carbânion, conforme apurado por *Chang* e *Perrin*<sup>117</sup>.



**Figura 9:** Diagrama de energia livre de Gibbs para a formação de chalcona catalisada por base, experimentalmente por *Chang* e *Perrin*<sup>117</sup>, e calculadas em ambientes aquoso e acetonitrila nos níveis de teoria B2PLYP/AUG-cc-pVDZ//B3LYP/6-31+G(3df,3dp). Nomenclatura das estruturas foram definidas na Figura 8.

**Tabela 2:** Frequências imaginárias ( $\text{cm}^{-1}$ ) e energias livre de Gibbs ( $\text{kcal mol}^{-1}$ ) as espécies da Figura 8 para diferentes solventes.

Espécies	$v_{H_2O}^\ddagger$	$H_2O$	$v_{D_2O}^\ddagger$	$D_2O$	$CH_3CN$	(i) <sup>113</sup>	(ii) <sup>117</sup>	(iii) <sup>122</sup>
<b>2</b>	—	0,00	—	0,00	0,00	0,00	0,00	0,00
<b>1‡</b>	1068	18,90	795	19,93	10,74	18,28	18,28	18,50
<b>1 + 2<sup>-</sup></b>	—	6,42	—	4,46	-6,41	5,78	5,78	—
<b>2‡</b>	254	22,49	252	20,37	11,63	17,55	19,02	20,90
<b>3<sup>-</sup></b>	—	6,94	—	4,86	-0,10	-0,33	0,31	—
<b>3‡</b>	374	8,55	289	5,88	1,16	6,21	2,17	—
<b>3</b>	—	9,09	—	7,37	1,87	0,86	0,01	—
<b>4‡</b>	1227	24,47	901	24,65	14,41	18,24	17,30	23,20
<b>5</b>	—	9,59	—	7,39	-0,96	3,48	4,53	—
<b>5‡</b>	170	22,42	189	19,01	17,60	21,47	20,12	19,80
<b>4</b>	—	0,76	—	0,66	0,19	-0,05	-1,54	—

i) Perfil completo da energia livre de *Gibbs* experimental de *Guthrie*<sup>113</sup>; ii) perfil completo da energia livre de *Gibbs* experimental de *Chang* e *Perrin*<sup>117</sup>; iii) perfil de energia livre de *Gibbs* teórico para formação de chalconas mediada por alcóxido metálico e auto catalisada por ácido benzílico<sup>122</sup>.

Pela Figura 9 é possível observar o efeito de solvente, decisivo nas sequências específicas da reação: (i) para as etapas **R1** e **R2**, o meio contínuo de água se ajusta melhor ao diagrama experimental de energia livre; e (ii) para as etapas **R3** a **R5**, o perfil em meio contínuo de acetonitrila está mais próximo do diagrama experimental. O comportamento conflitante em água e acetonitrila pode ser esclarecido com base na condição de que ambos os estudos cinéticos experimentais de formação de chalcona em água adicionaram quantidades consideráveis de acetonitrila ( $CH_3CN$ ) como cossolvente<sup>117,134</sup>, permitindo uma contribuição de solvente misto.

A formação de um poço no potencial de energia livre em equilíbrio com a etapa **R3** privilegia as espécies **3**<sup>-</sup>. Esta forma do perfil de energia livre está bem documentada na literatura para explicar solvatômeros em ácidos carboxílicos<sup>139</sup> e tautomerismo enol-enol<sup>140</sup>, e não está relacionada a polaridade do solvente, mas é inerente ao distúrbio molecular causado pelo solvente em torno da coordenada de reação de transferência de prótons<sup>136,140</sup>. Apesar de aplicar modelos de solvatação contínua, as correções entrópicas estatísticas de nossos cálculos foram adequadas para capturar este comportamento em ambos os solventes. Cálculos de dinâmicas moleculares quânticas em solvente explícito permitem avaliar grande diversidade conformacional em reações orgânicas e este protocolo foi aplicado para avaliar a condensação aldólica na fase aquosa de formaldeído e acetona<sup>123</sup>. Os efeitos conformacionais e de solvente apresentados nos cálculos identificaram canais de múltiplas etapas; no entanto, a topografia geral do mecanismo de *Claisen-Schmidt* foi recuperada.

### **1.6.3 Individualização de fatores físico-químicos na cinética da condensação de *Claisen-Schmidt***

O protocolo moderno para encontrar a Etapa Lenta de Controle (RCS, do inglês: *Rate Control Step*)<sup>141,142</sup> da cinética da reação é baseado na análise do perfil de energia livre completo, dividindo a reação em sequências onde os intermediários são mais estáveis do que os reagentes, procedendo até os produtos: a seção com mais diferença de energia leva à etapa de controle da constante cinética. A aplicação desta metodologia ao diagrama de energia livre proposto por Guthrie<sup>113</sup> e Chang-Perrin<sup>117</sup> para a condensação de *Claisen-Schmidt* (ver Tabela 2) confere a

**R5** a posição de RCS. O perfil calculado no meio contínuo de água confere a **R4** a posição da de RCS da reação em contraste com a proposição de Guthrie e Chang-Perrin; no entanto, em meio contínuo de acetonitrila ( $CH_3CN$ ), **R5** é recuperado como a RCS da reação.

O efeito isotópico do solvente em água também foi avaliado e comparados aos casos em que **R4** e **R5** foram considerados como RCS da reação. O parâmetro  $iKIE = k_{obs}(D_2O)/k_{obs}(H_2O)$  foi definido para facilitar a interpretação do efeito isotópico cinético inverso, onde o numerador representa a constante cinética da reação na água pesada ( $D_2O$ ) e o denominador na água ( $H_2O$ ). A cinética de seqüências específicas da reação também foi avaliada, o que leva: (i) **3** a **1** e **2**; (ii) **3** a **4**; e (iii) **1** e **2** a **3**.

Detalhes acerca da hipótese cinética e composição de equações para o tratamento das constantes cinéticas podem ser encontradas na publicação deste trabalho<sup>108</sup>.

A partir das constantes cinéticas das reações elementares diretas e inversa e aplicando a teoria do estado de transição com a correção de *Kramer*, foram obtidos os valores de  $k_{obs}(H_2O) = 0,0023 M^{-2} s^{-1}$  e  $iKIE = 27,61$  à temperatura de  $298K$ , resultados semelhantes aos obtidos quando **R4** é assumido como a RCS da reação (ver Tabela 2). Quando **R5** é assumido como a RCS da reação, o valor obtido foi  $k_{obs}(H_2O) = 0,0145 M^{-2} s^{-1}$ , muito próximo ao obtido por Chang e Perrin<sup>117</sup> em uma mistura de água e acetonitrila como solventes, que foi de  $0,0111 M^{-2} s^{-1}$ . No entanto, o parâmetro  $iKIE = 2028$  apresenta uma discordância completa quando comparado ao valor experimental, que é de 1,14. Quando a correção de tunelamento de Skodje e Truhlar é adicionada considerando os passos **R4** e **R5** como RCS's, as constantes cinéticas fornecem valores próximos:  $0,0365 M^{-2} s^{-1}$  e  $0,0305 M^{-2} s^{-1}$  respectivamente. As constantes cinéticas aproximam-se dos resultados obtidos experimentalmente com solvente próticos polares (soluções de etanol). É importante observar que o efeito isotópico inverso permanece mesmo assumindo **R4** como a RCS da reação, comportamento obtido devido ao  $iKIE > 1$  na etapa **R2**. O valor calculado para a constante cinética da reação global foi de  $k_{obs}(H_2O) = 0,0166 M^{-2} s^{-1}$ , recuperando o valor obtido experimentalmente por Chang e Perrin<sup>117</sup> e produzindo um valor de  $iKIE = 10,12$  em acordo semiquantitativo com o valor experimental.

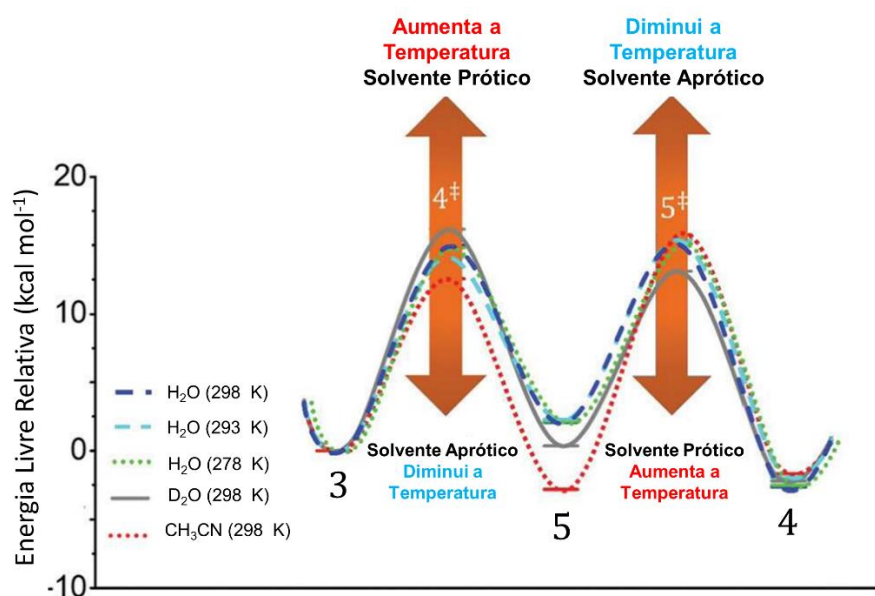


Cálculos realizados demonstraram que, partindo do cetol **3**, as constantes cinéticas para a formação dos reagentes **1** e **2** foram muito maiores do que para formação da chalcona **4** (ver Tabela 2). Estes resultados estão de comum acordo com a hipótese de *Chang-Perrin* que exclui as etapas **R1**, **R2** e **R3** como RCS's da reação. Nossos cálculos mostraram que na ausência da presença da correção de tunelamento não é possível observar um efeito isotópico inverso na cinética em  $k_{3 \rightarrow 1+2}$ ,  $k_{3 \rightarrow 4}$  e  $k_{1+2 \rightarrow 3}$ , uma vez que os efeitos do isótopo nas altas barreiras intrínsecas das etapas elementares são negligenciáveis. Sob o ponto de vista das correções de tunelamento quântico, apenas as etapas **R2** e **R5** apresentaram diminuição de suas barreiras intrínsecas para os casos deuterados. Enquanto em **R2** foi uma pequena variação, em torno de  $1,30 \text{ kcal mol}^{-1}$ , para **R5** foi desprezível,  $0,25 \text{ kcal mol}^{-1}$ . É digno de nota que os efeitos de tunelamento implicam proximidade das constantes cinéticas das reações elementares das etapas **R4** e **R5**, tornando inútil a aplicação do conceito de RCS (as condições para aplicação deste modelo foram estudadas por Laidler<sup>143</sup>).

Só foi possível reproduzir qualitativamente as relações entre  $k_{3 \rightarrow 1+2}$ ,  $k_{3 \rightarrow 4}$  e  $k_{1+2 \rightarrow 3}$ , ver na Tabela 2. Este comportamento pode ser esclarecido considerando que a topografia completa do perfil de energia livre foi obtida em meio aquoso contínuo. No entanto, a presença de acetonitrila como cossolvente no procedimento experimental pode levar a uma variação apreciável nas barreiras intrínsecas das etapas **R4** e **R5**, conforme atestado anteriormente (ver Figura 9 e Figura 10). A temperatura também pode levar a pequenas mudanças nas barreiras intrínsecas das etapas **R4** e **R5**: com uma diminuição na temperatura de 298 para 293 e 273K, os valores de constante cinética obtidos sugerem a etapa **R5** como a RCS (consultar Tabela 2 e Figura 10). Este comportamento é responsável pela tendência da barreira intrínseca na etapa **R4** para diminuir sistematicamente enquanto da etapa **R5** aumenta. Na Figura 10, a energia livre de *Gibbs* foi recalculada a partir de constantes cinéticas das reações elementares considerando todos os fatores físico-químicos: solvente contínuo, efeito isotópico, temperatura e contribuições dinâmicas.

O arquivo *Support Information*, fornecido junto a publicação deste trabalho<sup>108</sup> mostra os ajustes das constantes cinéticas das reações elementares da Figura 9 usando a lei fenomenológica Aquilanti-Mundim (mais informações nos trabalhos de Aquilanti e colaboradores<sup>30</sup>; e Carvalho-Silva e colaboradores<sup>84</sup> e referências nelas

contidas). Com as equações ajustadas na faixa de temperatura de 273 a 393K é possível extrapolar o efeito da temperatura em um ambiente aquoso no diagrama de energia livre da condensação de *Claisen-Schmidt*. Como já discutido previamente, a lei de Aquilanti-Mundim pode correlacionar a curvatura no gráfico de *Arrhenius* com efeitos quânticos (sub-*Arrhenius*) ou fenômenos de transporte clássicos (super-*Arrhenius*). A curvatura da dependência da temperatura das constantes cinéticas nas reações elementares da condensação de *Claisen-Schmidt* enfatiza esses comportamentos, com as etapas **R1** e **R4** influenciadas por efeitos de tunelamento quântico (côncavo)<sup>61,144</sup> e as etapas **R2**, **R3** e **R5** por efeitos de fricção e difusão clássicas do solvente (convexo)<sup>23,145</sup>.



**Figura 10:** Diagrama de energia livre de Gibbs para as etapas R4 e R5 enfatizando os efeitos: isotópicos, de temperatura e de solvente contínuo.

#### 1.6.4 Relação com o trabalho de Chang e Perrin

Em resumo, nossos cálculos corroboraram a conclusão de Chang e Perrin apoiando a reversão mais rápida do cetol **3** para os reagentes **1** e **2** quando comparado com a progressão para **4**. No entanto, expandimos a compreensão dos efeitos do solvente e da temperatura na seleção de **R4** ou **R5** como a RCS para a formação de chalcona: um aumento na temperatura e a presença de um solvente prótico induz o **R4** como o RSC. Além disso, a hipótese de que o efeito isotópico na cinética inversa da reação global eliminou a etapa **R4** como sendo a RCS deve ser analisada cuidadosamente, uma vez que nossos resultados demonstraram que a

etapa **R2** também fornece esta condição limitante: efeitos isotópicos na cinética experimental de  $k_{1+2\rightarrow 3}$  e  $k_{\rightarrow 4}$  podem elucidar este mecanismo.

Este trabalho foi publicado em janeiro de 2021<sup>108</sup> com o título: “*Topography of the free energy landscape of Claisen–Schmidt condensation: solvent and temperature effects on the rate-controlling step*”. Logo após, em abril de 2021, o professor Charles Perrin submeteu para publicação um outro artigo<sup>146</sup> intitulado: “*Comment on ‘Topography of the free energy landscape of Claisen–Schmidt condensation: solvent and temperature effects on the rate-controlling step’ by N. D. Coutinho, H. G. Machado, V. H. Carvalho-Silva and W. A. da Silva, Phys. Chem. Chem. Phys., 2021, 23, 6738*”, no qual ele discorda da nossa hipótese e reafirma a existência de uma RCS no mecanismo reacional. Entretanto, como já foi extensamente discutido nesta seção, o mecanismo de condensação de *Claisen–Schmidt* é fortemente afetado por diferentes condições físico-químicas, provocando mudanças substanciais no perfil de energia livre de *Gibbs* das etapas **R1** a **R5**. As barreiras de energia livre calculadas por nós e obtidas experimentalmente por Chang e Perrin<sup>117</sup>; e Guthrie e colaboradores<sup>134</sup> para **R1**, **R2**, **R4** e **R5**, mostraram magnitudes muito semelhantes entre as etapas citadas (ver Tabela 2) e, portanto, uma análise cinética baseada apenas em identificar a RCS pode causar um mal-entendido. Contudo, uma réplica ao comentário do professor Charles Perrin foi por nós submetida para publicação em agosto de 2021<sup>147</sup>, intitulada “*Reply to the ‘Comment on “Topography of the Free Energy Landscape on the Claisen–Schmidt Condensation: Solvent and Temperature Effect in the Rate-Controlling Step” by N. D. Coutinho, H. G. Machado, V. H. Carvalho-Silva and W. A. da Silva, Phys. Chem. Chem. Phys., 2021, 23, 6738*”, que foi prontamente aceita para publicação, inclusive sendo capa da edição naquela ocasião. Nesta publicação é discutido os motivos que levaram às falhas de compreensão do mecanismo de síntese, principalmente as limitações experimentais relacionadas à necessidade de consideráveis adições de acetonitrila como um cossolvente a fim de manter a solubilidade dos substratos e do produto chalcona.

## CAPÍTULO 2

### $Q_{d10}$ – GSA

#### 2.1 Introdução

O coeficiente  $Q_{10}$ <sup>148–150</sup> foi amplamente descrito na literatura como uma medida do grau de dependência da temperatura pelo tempo de ocorrência de um determinado processo químico ou biológico. No entanto, as conclusões tiradas por diferentes pesquisadores com o uso desse coeficiente não são universais<sup>150,151</sup> e nem sempre são consistentes entre si<sup>152–169</sup>. Portanto, é razoável presumir que este problema decorre do fato de que o coeficiente  $Q_{10}$  provavelmente não deve ser definido pela razão entre as constantes cinéticas definidas pela lei de *Arrhenius*, pois nestes casos um aumento de temperatura em 10 unidades não acarreta a duplicação ou a triplicação da constante cinética de um determinado processo, como geralmente é descrito na literatura<sup>150</sup>.

Com base nesta hipótese, nosso grupo propôs o modelo de coeficiente generalizado de temperatura  $Q_{d10}$ <sup>32</sup>, que unifica os resultados conflitantes obtidos pelos pesquisadores. Para incentivar o uso deste modelo entre a comunidade científica, foi desenvolvido um código com uma interface gráfica intuitiva e amigável ao usuário, que pode realizar diferentes procedimentos para calcular o  $Q_{10}$  e outras propriedades dependentes da temperatura. A novidade deste código é seu escopo geral e exploração particular das formulações  $Q_{d10}$  para reproduzir os comportamentos *Arrhenius* e não-*Arrhenius* em diferentes temperaturas, um tópico que tem sido foco de investigações recentes em processos biológicos. O código é referido como  $Q_{d10}$  – GSA e opera com conceitos relacionados ao processo de difusão e constantes cinéticas, energia de ativação e coeficiente  $Q_{10}$ . Pode ser usado para estimar parâmetros fenomenológicos dos gráficos de *Arrhenius* e *d-Arrhenius* usando o método de otimização estocástica GSA<sup>7</sup>.

O  $Q_{d10}$  – GSA apresenta uma interface gráfica intuitiva e amigável, acreditamos que possa servir como uma excelente ferramenta de trabalho para caracterização de dados experimentais, além de enfatizar a necessidade de generalizar a abordagem usual  $Q_{10}$  e servir como material didático para o ensino do

conceito da dependência da velocidade de processos com a temperatura. Deste modo, neste Capítulo 2 será descrito todo o procedimento de cálculo para o modelo de coeficiente generalizado  $Q_{d10}$  e demonstrado, a partir de aplicações utilizando o *software*  $Q_{d10}$  – GSA, que as constantes cinéticas de processos químicos e biológicos não obedecem às leis exponenciais (lei de *Arrhenius*) e que uma revisão da definição padrão do coeficiente  $Q_{10}$  foi necessária.

Ao final do texto na seção de apêndice B serão abordadas questões relativas à obtenção, instalação e dicas para utilização do código desenvolvido.

## 2.2 Coeficiente da temperatura $Q_{10}$

### *O modelo usual de Arrhenius*

O modelo usual para o coeficiente  $Q_{10}$  é expresso pela Equação (36):

$$Q_{10} = \left[ \frac{k(T_2)}{k(T_1)} \right]^{10/(T_2-T_1)} \quad (36)$$

na qual  $T$  é a temperatura em graus Celsius ou Kelvin; e  $k$  é a constante cinética expressa como um decaimento exponencial da temperatura, a equação de *Arrhenius*, Equação (2).

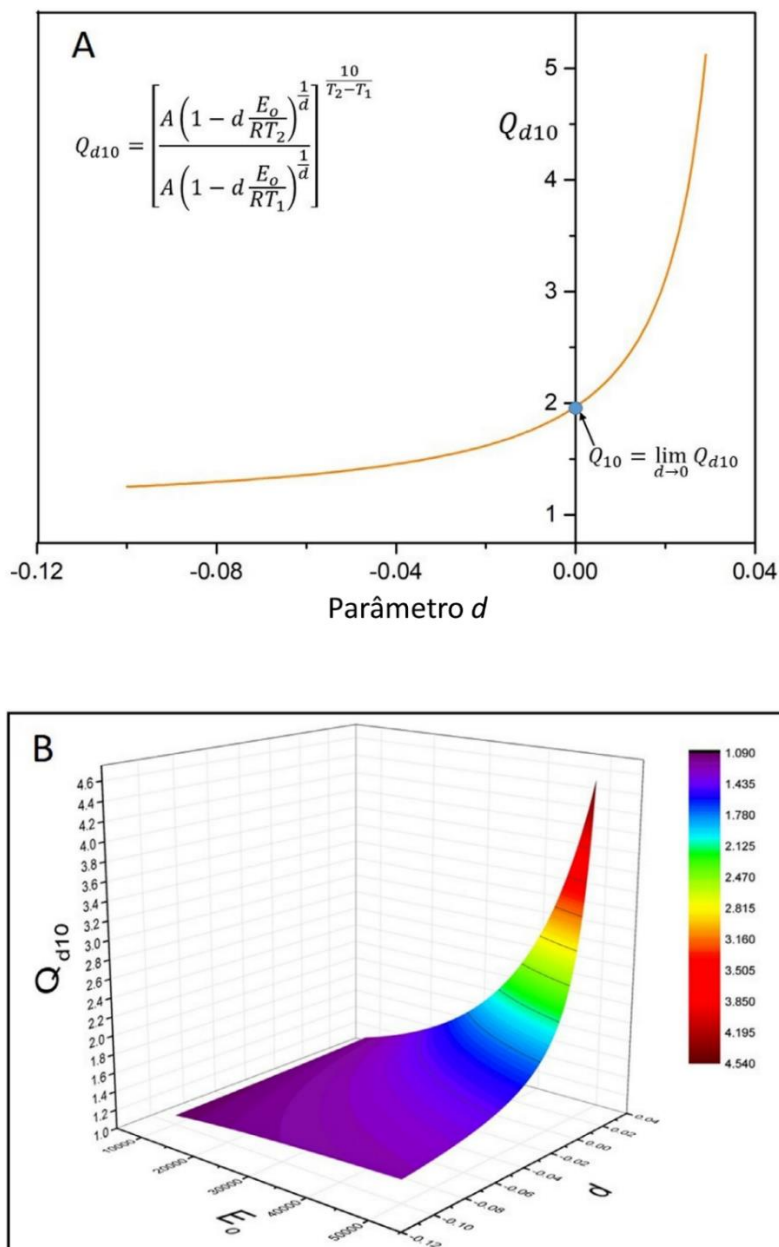
### *O Modelo de Arrhenius deformado*

O modelo matemático proposto é baseado na teoria das funções exponenciais deformada na qual a constante cinética pode ser expressa pela formulação de Aquilanti-Mundim, Equação (12), na qual a energia de ativação é expressa na Equação (34).

Utilizando as Equações (36) e (12), o novo coeficiente  $Q_{d10}$  é descrito na Equação (37).

$$Q_{d10} = \left[ \frac{k_d(T_2)}{k_d(T_1)} \right]^{10/(T_2-T_1)} \quad (37)$$

Ao aplicar nas Equações deformadas (12) e (37) o limite com  $d \rightarrow 0$  verifica-se que ambas recuperam suas formas clássicas não deformadas, as Equações (2) e (36):  $\lim_{d \rightarrow 0} k_d(T) = k(T)$ ; e  $\lim_{d \rightarrow 0} Q_{d10} = Q_{10}$ . Portanto, podemos afirmar que os modelos deformados são gerais enquanto as equações clássicas representam casos particulares desses modelos (ver Figura 11).



**Figura 11:** Coeficiente de temperatura  $Q_{10}$  generalizado. (A) Comportamento do coeficiente  $Q_{d10}$  como uma função do parâmetro de deformação  $d$ , assumindo a energia de ativação  $E_0$  como uma constante. Em casos extremos nos quais  $d = 0$ ,  $k_{T+10}/k_T \cong 2$  e a o modelo padrão é recuperado. (B) Comportamento do coeficiente  $Q_{d10}$  de acordo com o parâmetro de deformação  $d$  e a energia de ativação  $E_0$ .

O sinal do parâmetro  $d$  controla o comportamento do funcional da lei deformada, sendo responsável por reproduzir os desvios da linearidade que ocorrem no gráfico de Arrhenius:  $d < 0$  reproduz o comportamento *sub-Arrhenius* no qual o gráfico é côncavo;  $d > 0$  reproduz o comportamento *super-Arrhenius* no qual o gráfico é convexo; e  $d = 0$  representa a forma linear convencional de Arrhenius. A Figura 11A mostra o comportamento do coeficiente  $Q_{d10}$  como uma função do parâmetro de deformação, assumindo a energia de ativação  $E_0$  como constante.

No caso do limite utilizando  $d \rightarrow 0$ , temos que  $k_{T+10}/k_T \cong 2$ , isto é, a constante cinética dobra seu valor a cada variação de 10 unidades da temperatura, como previsto pela teoria usual ( $Q_{10}$ ). Entretanto, para casos gerais onde  $d \neq 0$  (comportamentos não-Arrhenius, ver Figura 11) a razão entre as constantes cinéticas pode assumir qualquer valor<sup>170–173</sup>. Neste novo modelo, os valores de  $Q_{d10}$  são também função da energia de ativação  $E_0$ , como mostrado na Figura 11B.

## 2.3 Exemplos e aplicações

O coeficiente  $Q_{10}$  é amplamente utilizado para descrever uma variedade de fenômenos, incluindo: cinética química<sup>30</sup>; respiração em organismos vivos<sup>153,159,174–181</sup>; desempenho e velocidade da contração das fibras musculares<sup>154,155,162,182–184</sup>; fibrilação atrial<sup>185,186</sup>; degradação do carbono do solo<sup>187–190</sup>; efeito de temperatura nos canais iônicos<sup>156,164,191–193</sup>; e taxas metabólicas<sup>152,161,166,167,169,194–198</sup>; mas em casos que o gráfico de *Arrhenius* não é válido, os valores do coeficiente  $Q_{10}$  não são consistentes com os dados experimentais.

Será aqui demonstrado que a teoria proposta está de acordo com os dados experimentais, principalmente em faixas de temperatura que impossibilitam o uso da formulação comum. Os casos abordados no presente estudo (descritos nas próximas sessões) envolvem o estudo do efeito da temperatura nos processos metabólicos em organismos vivos, decomposição do carbono no solo, processos fisiológicos e consumo de oxigênio e outras substâncias químicas por plantas e animais. Gráficos de *Arrhenius* deformados e energia de ativação (em  $kJ/mol$ ) foram usados para padronizar os casos distintos abordados neste estudo.

### 2.3.1 Metabolismo do Zooplâncton do Ártico

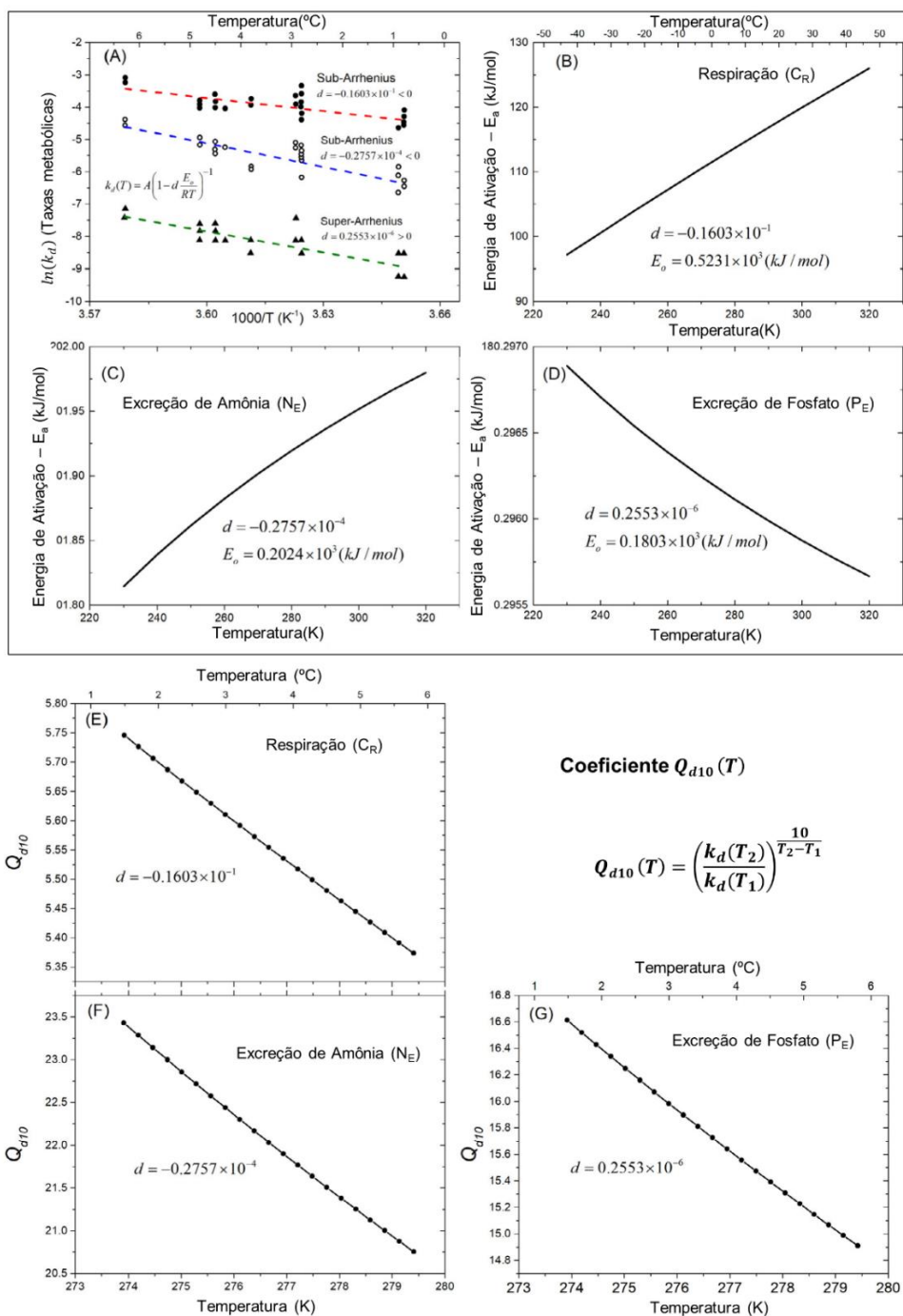
Para a primeira aplicação do modelo proposto foram analisados os resultados obtidos por Alcaraz e colaboradores<sup>153</sup>. Neste estudo os autores investigaram o efeito da temperatura no metabolismo do zooplâncton do Ártico. Com base nesses dados experimentais, as taxas deformadas ( $k_d$ ) foram ajustadas para calcular o coeficiente  $Q_{d10}$  e a energia de ativação em diferentes temperaturas. Os resultados são mostrados na Figura 12A. Mais especificamente, foi avaliado a dependência da temperatura com as taxas de respiração e de excreção de fosfato e amônia. Nosso

modelo retornou um comportamento *sub-Arrhenius* ( $d < 0$ ) associado às taxas de respiração e excreção de amônia, enquanto a taxa de excreção de fosfato teve um comportamento ligeiramente *super-Arrhenius* ( $d > 0$ ), mas próximo de zero.

Como os valores dos parâmetros de deformação são diferentes de zero, a energia de ativação ( $E_a$ ) das reações depende da temperatura (Figura 12B-D), ao contrário do caso comum em que a  $E_a$  é considerada constante. Para os processos de respiração e excreção de amônia a energia de ativação aumentou com o aumento da temperatura (Figura 12B e D), enquanto na excreção de fosfato o cenário foi invertido: a energia de ativação diminuiu com o aumento da temperatura (comportamento *super-Arrhenius*, Figura 12D). Ao contrário do processo respiratório, a energia de ativação nos processos de excreção tornou-se praticamente constante em temperaturas superiores a  $300K$  e mudou abruptamente em baixas temperaturas. Consequentemente, os coeficientes  $Q_{d10}$  para as três taxas diminuíram com o aumento da temperatura de forma não exponencial ( $d \neq 0$ ), conforme mostrado na Figura 12A. Esses resultados demonstraram que as taxas metabólicas do zooplâncton do Ártico exibem uma relação não linear com a temperatura na perspectiva do gráfico de *Arrhenius*. Além disso, o metabolismo do zooplâncton ártico é altamente sensível às mudanças de temperatura ( $d \neq 0$ ) tanto para respiração quanto para excreção. Assim, respostas abruptas não lineares, mudanças de regime e trajetórias complexas são esperadas durante a evolução dessas espécies.

O valor da energia de ativação no caso da respiração (Figura 12B) não foi constante, indicando uma alta sensibilidade do processo respiratório às mudanças de temperatura. Em contraste, nos processos de excreção (Figura 12C e D), a energia de ativação foi fracamente dependente da temperatura ( $d \cong 0$ ). No caso da amônia (NE), a excreção aumentou com o aumento da temperatura ( $d < 0$  ou *sub-Arrhenius*), enquanto no caso do fosfato, a energia de ativação diminuiu com o aumento da temperatura ( $d > 0$  ou *super-Arrhenius*). Na pequena faixa de temperaturas ( $273K$  a  $280K$ ) usada na análise, a sensibilidade à temperatura dada pelo parâmetro  $Q_{d10}$  exibiu um comportamento quase linear nos casos de respiração ( $C_R$ ), excreção de amônia ( $N_E$ ) e excreção de fosfato ( $P_E$ ), como mostrado na Figura 12E-F.





**Figura 12:** Efeito da temperatura no metabolismo do zooplâncton do ártico pelo modelo de coeficiente deformado ( $Q_{d10}$ ). (A) Gráficos de Arrhenius deformados nos casos de respiração ( $C_R$ ) e taxas de excreção de amônia ( $N_E$ ) e fosfato ( $P_E$ ). Energia de ativação, usando d-Arrhenius para as taxas de (B) respiração ( $C_R$ ), (C) excreção de amônia ( $N_E$ ) e (D) excreção de fosfato ( $P_E$ ), em ( $C_R$ ,  $N_E$  e  $P_E$ )  $\mu mol C_{ZOO}^{-1} day^{-1}$ . Coeficiente deformado ( $Q_{d10}$ ) para as taxas de (E) respiração ( $C_R$ ), (F) excreção de amônia ( $N_E$ ) e (G) excreção de fosfato ( $P_E$ ).

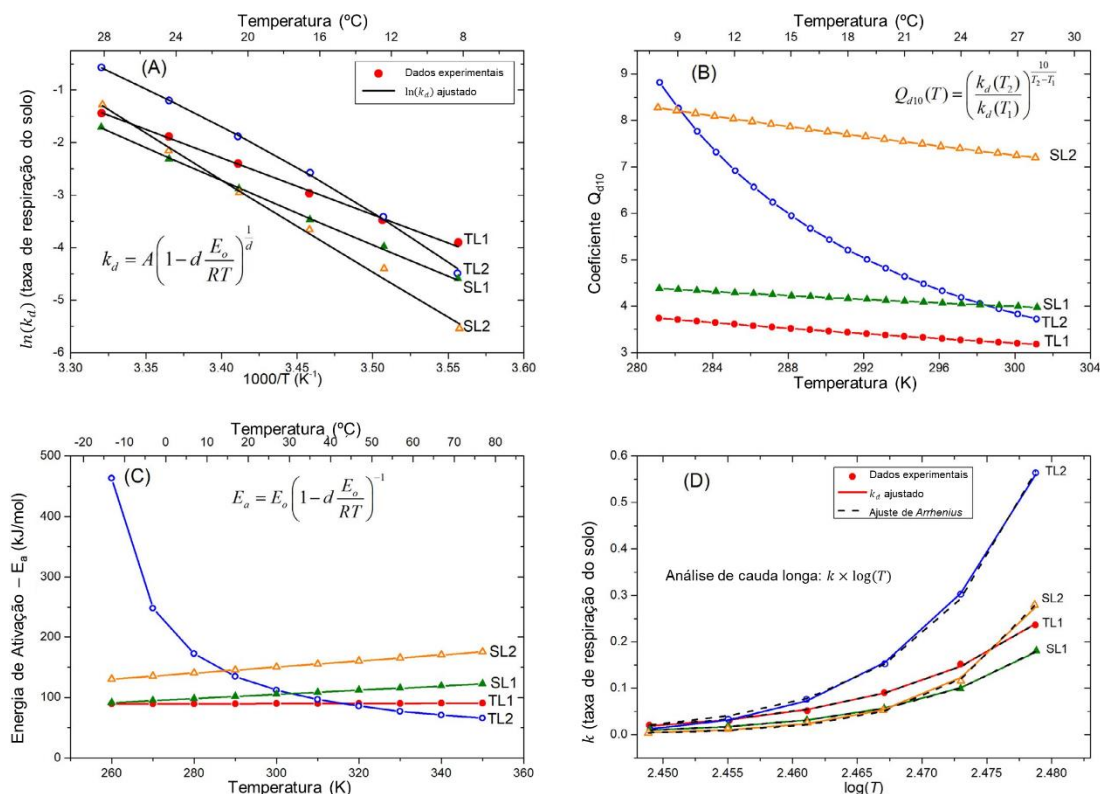
### 2.3.2 Processos de Decomposição de Carbono Orgânico do Solo

A compreensão acerca da decomposição do Carbono Orgânico do Solo (SOC, do inglês: *Soil Organic Carbon*) é essencial para prever o impacto das mudanças climáticas nos ecossistemas terrestres: o SOC contém aproximadamente três vezes mais carbono que a vegetação terrestre e duas vezes mais que a atmosfera, visto que o aquecimento global já afetou visivelmente o meio ambiente<sup>189</sup>.

Vários estudos<sup>163,165,168,187-189,199-203</sup> levantaram recentemente questões sobre a confiabilidade dos métodos usuais para caracterizar a sensibilidade à temperatura de processos que ocorrem durante a decomposição do SOC. Atualmente, o coeficiente  $Q_{10}$  é frequentemente usado como uma medida para inferir o efeito da temperatura na decomposição do SOC. Entretanto, considerando a decomposição do SOC como um problema multifatorial e focando apenas na influência da temperatura sobre este fenômeno, o coeficiente  $Q_{10}$  em sua forma usual não é um parâmetro de medição adequado, pois é definido como a razão entre as taxas que são assumidas, a priori, como sendo funções exponenciais da temperatura e, portanto, o gráfico de *Arrhenius* deveria ser linear, o que contradiz os dados experimentais obtidos por Yan<sup>189</sup> e Tang<sup>188</sup>.

Um dos principais fatores que causam esta contradição é a sensibilidade à temperatura, fato observado nos casos que serão aqui tratados. Portanto, no sentido de contribuir para um melhor entendimento deste problema, os dados experimentais foram revisados e exigiram uma reformulação do método do cálculo do coeficiente  $Q_{10}$  para o modelo generalizado  $Q_{d10}$ . Os resultados são mostrados na Figura 13.

Nos estudos de Yan<sup>189</sup> e Tang<sup>188</sup> foram estimadas as diferenças na decomposição do SOC e na sensibilidade à temperatura entre diferentes amostras de solo em áreas florestadas, bem como entre solos superficiais e camadas de subsolo em sistemas de cultivo, utilizando o modelo  $Q_{10}$ . A aplicação do modelo proposto à decomposição SOC mostrou que as taxas de respiração do solo analisadas por Yan<sup>189</sup> não variaram exponencialmente com a temperatura (ver Figura 13D). Isso implica um desvio do modelo de *Arrhenius* usual porque os coeficientes de deformação ( $d$ ) não foram zero em todos os sistemas avaliados no estudo.



**Figura 13:** Sensibilidade à temperatura de processos de decomposição do SOC através do modelo de coeficiente deformado ( $Q_{d10}$ ). Influência da temperatura na taxa de decomposição do SOC em quatro sistemas diferentes: TL1: camada superficial do solo do arrozal – local 1; TL2: camada superior do solo do planalto - local 2; SL1: camada de subsolo do arrozal – local 1; e SL2: camada de subsolo do planalto – local 2. (A) Taxas de decomposição experimental (pontos) com o aumento da temperatura e o ajuste (linhas sólidas) usando o modelo apresentado na Equação (12); (B) coeficientes  $Q_{d10}$  versus temperatura; (C) comportamento da energia de ativação com o aumento da temperatura.

Visivelmente, no caso de TL2 (camada superficial do solo 2), o processo foi do tipo *super-Arrhenius*, ou convexo ( $d > 0$ ) (Figura 13A e B), e a energia de ativação foi altamente sensível às variações de temperatura, conforme mostrado na Figura 13C. Nesse processo, o comportamento do novo coeficiente  $Q_{d10}$  indica uma forte influência da temperatura na decomposição do SOC. Este caso é um exemplo clássico em que a definição padrão do coeficiente  $Q_{10}$  não se aplica.

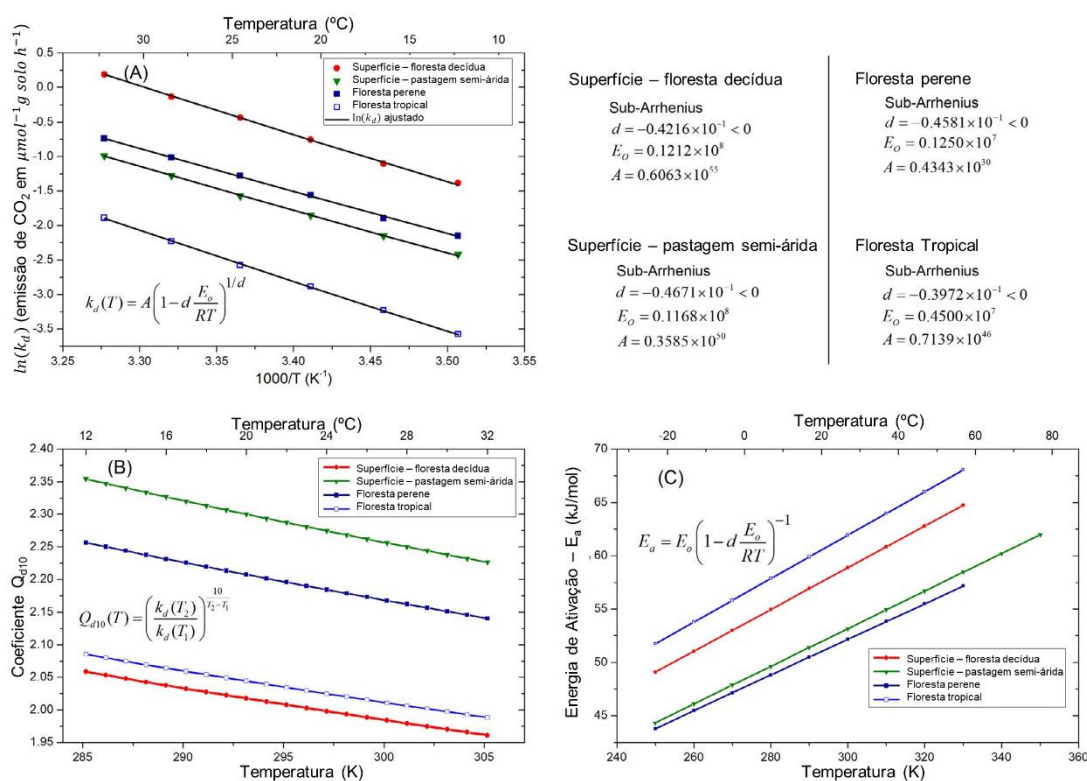
Os outros três casos (TL1, SL1 e SL2) exibiram comportamento *sub-Arrhenius* quase linear ( $d \cong 0$ ). Esses resultados indicaram que a taxa respiratória (SOC) na camada superficial do solo (TL) é mais sensível às variações de temperatura do que na camada subsolo (SL). O coeficiente  $Q_{d10}$  em casos de superfície (TL) estabelece uma razão de taxa de  $3.5 < \frac{k_d(T_2)}{k_d(T_1)} < 9.0$  entre 7 e 30°C. Os parâmetros de ajuste são mostrados na Tabela 3.

**Tabela 3:** Parâmetros usados no ajuste do modelo deformado de Arrhenius para os dados experimentais.

Tipo de Solo	TL1	TL2	SL1	SL2
Tipo de taxa	Sub-Arrhenius	Super-Arrhenius	Sub-Arrhenius	Sub-Arrhenius
$d$	$-0,1014 \times 10^{-2}$	0,1095	$-0,2336 \times 10^{-1}$	$-0,1645 \times 10^{-1}$
$E_o(kJ/mol)$	$0,9338 \times 10^2$	$0,1893 \times 10^2$	$0,8091 \times 10^4$	$0,2627 \times 10^5$
$A$	$0,1892 \times 10^{16}$	$0,5382 \times 10^7$	$0,7464 \times 10^{80}$	$0,3636 \times 10^{136}$

### 2.3.3 Processos de Decomposição de SOC em diferentes tipos de solo

No mesmo contexto da seção anterior, aqui os autores investigaram a sensibilidade à temperatura da decomposição da SOC em diferentes tipos de solo (superfície de floresta decídua, superfície de pastagem semi-árida, floresta perenifólia e floresta tropical, ver Figura 14)<sup>188</sup>.



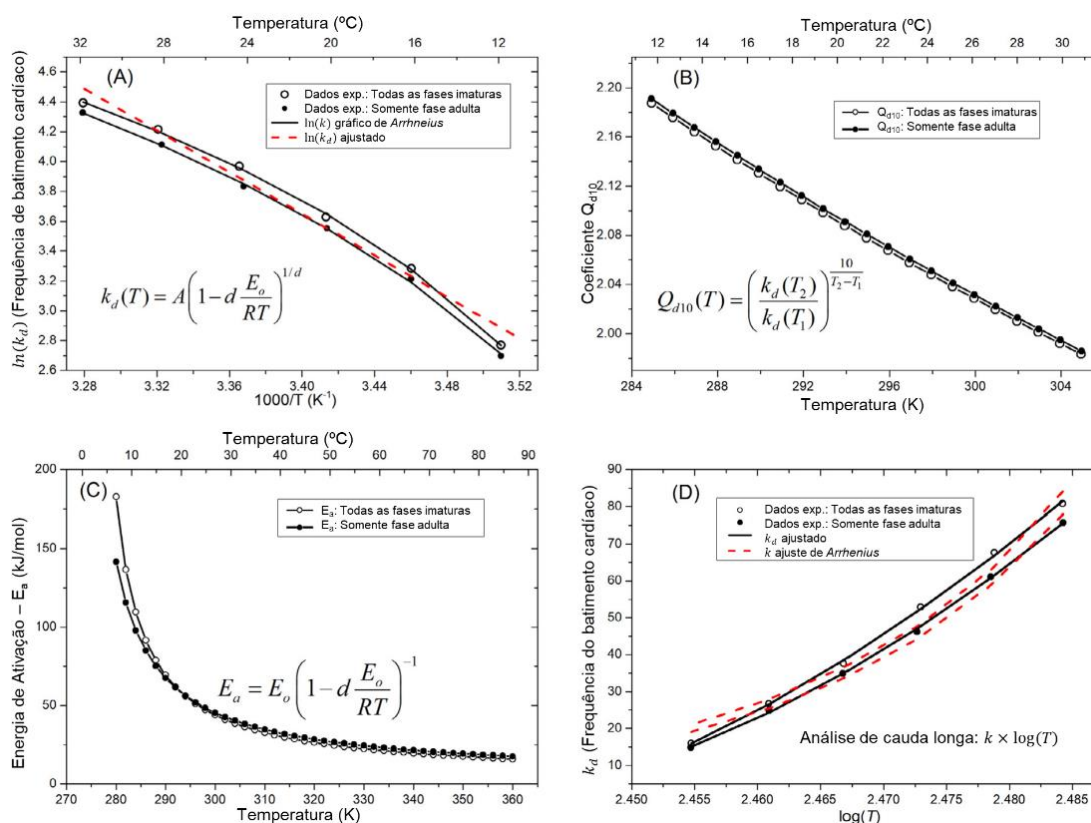
**Figura 14:** Influência da temperatura na taxa de decomposição do SOC em quatro sistemas diferentes: superfície de floresta estacional decidual, superfície de pastagem semiárida, floresta perene e floresta tropical. (A) Taxas de decomposição experimental (pontos) com o aumento da temperatura e o ajuste (linhas sólidas) usando o modelo apresentado na Equação (12); (B) coeficientes  $Q_{d10}$  versus temperatura; (C) variação da energia de ativação com a temperatura.

Revisitando esses resultados experimentais dentro do modelo generalizado  $Q_{d10}$ , constatou-se os coeficientes de deformação  $d < 0$  em todos os casos e, portanto, caracterizados como sendo do tipo *sub-Arrhenius* (Figura 14A-C). Os

coeficientes  $Q_{d10}$  obtidos levaram a razões de taxa de  $1.95 < \frac{k_d(T_2)}{k_d(T_1)} < 2.4$  entre 12 e  $32^\circ\text{C}$  (Figura 14B). Essas variações nas taxas de decomposição induzem uma mudança na energia de ativação para ser dependente da temperatura, ao contrário do modelo padrão que assume a energia de ativação como sendo constante.

### 2.3.4 Reações Fisiológicas dos Sistema Respiratório e Cardiovascular

As taxas de reação fisiológica, particularmente aquelas dos sistemas respiratório e cardiovascular, são fortemente afetadas pela temperatura corporal<sup>177,181,184</sup>. No presente estudo, a influência da temperatura na frequência cardíaca investigada na barata *Gromphadorhina portentosa* por Streicher e colaboradores<sup>177</sup> foi reavaliada utilizando o novo modelo de taxas aqui proposto (ver Figura 15).



**Figura 15:** Sensibilidade à temperatura nas reações fisiológicas dos sistemas respiratório e cardiovascular através do modelo de coeficiente deformado ( $Q_{d10}$ ). (A) Os resultados experimentais e o modelo proposto exibiram um comportamento convexo do gráfico  $\ln(k) \times 1/T$  o que indicou um desvio da teoria de Arrhenius padrão, indicado por uma linha tracejada. Os resultados experimentais são indicados por pontos sólidos e o modelo proposto é indicado por uma linha sólida; (B) em ambos os casos, os coeficientes  $Q_{d10}$  produziram uma razão de taxas de  $k_d(T_2)/k_d(T_1) \approx 2$ ; (C) a energia de ativação muda abruptamente com a temperatura; (D) os processos analisados foram, neste caso, do tipo super-Arrhenius ( $d > 0$ ).

As frequências cardíacas entre esta espécie não variam exponencialmente com a temperatura e, portanto, o gráfico de *Arrhenius* não se aplica neste caso. A Figura 15 mostra o comportamento da frequência respiratória (A), o coeficiente  $Q_{d10}$  (B) e a energia de ativação (C). As reações analisadas, neste caso, podem ser classificadas como do tipo *super-Arrhenius* ( $d > 0$ ) e a energia de ativação foi altamente sensível à variação da temperatura. Os coeficientes  $Q_{d10}$  exibiram uma fraca dependência não linear da temperatura corporal. A frequência cardíaca foi menos sensível às mudanças de temperatura na fase adulta do que nas fases imaturas. Isso ocorre porque o parâmetro de deformação  $d$  é menor no estágio adulto do que em outras fases do desenvolvimento. O coeficiente  $Q_{d10}$  em ambos os casos rendeu uma razão de taxas de  $2 < \frac{k_d(T_2)}{k_d(T_1)} < 2.2$  entre 280 e 305K. A variação da taxa com a temperatura (Figura 15D) indicou um comportamento *super-Arrhenius* ( $d > 0$ ). Os parâmetros de ajuste são mostrados na Tabela 4.

**Tabela 4:** Sistemas respiratório e cardiovascular Parâmetros usados no modelo deformado de *Arrhenius* como consequência do ajuste do  $Q_{d10}$  a partir dos dados experimentais.

Fases de maturação	Todas as Fases Imaturas	Somente Fase Adulta
Tipo de taxa	<i>Super-Arrhenius</i>	<i>Super-Arrhenius</i>
$d$	0,6050	0,5231
$E_o(kJ/mol)$	$0,3769 \times 10^4$	$0,4314 \times 10^4$
$A$	$0,3617 \times 10^4$	$0,5154 \times 10^4$

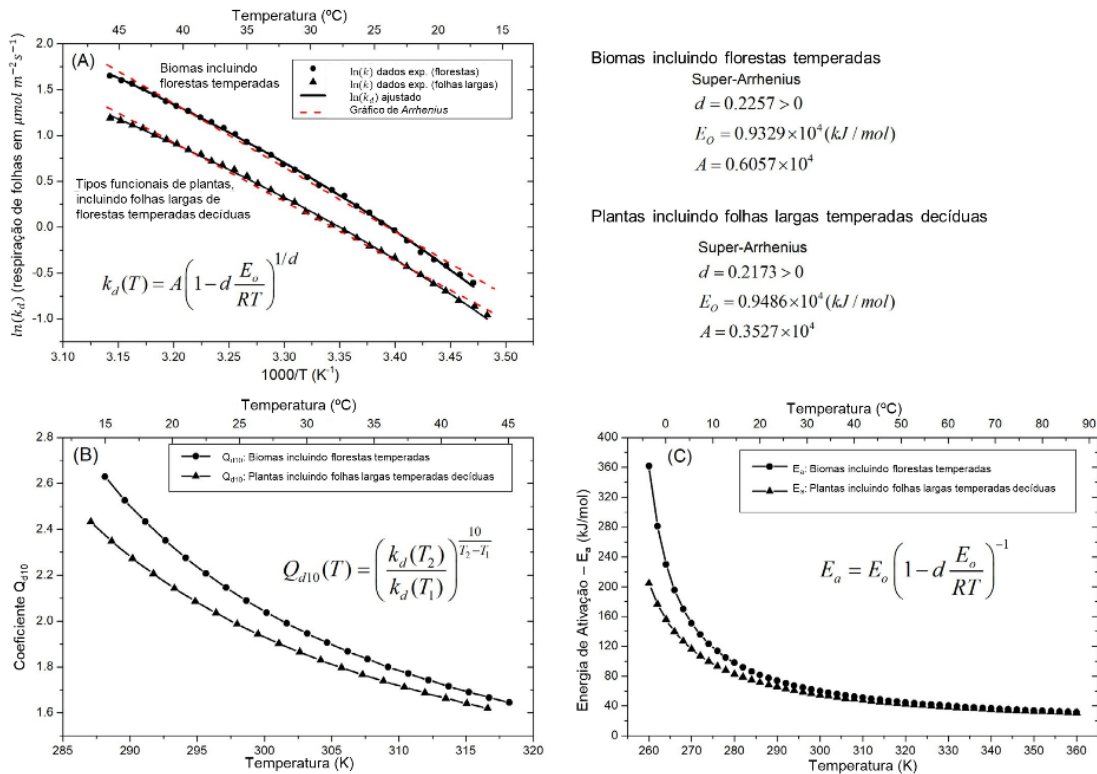
### 2.3.5 Respiração Foliar entre Biomas e tipos Funcionais de Plantas

A respiração das plantas envolve um fluxo maciço de carbono para a atmosfera e um grande controle sobre o ciclo global do carbono. *Heskel* e colaboradores<sup>180</sup> estudaram recentemente a respiração em 231 espécies de plantas em sete biomas e concluíram que os modelos exponenciais das taxas de respiração  $k(T)$  não descrevem corretamente a sensibilidade à temperatura de  $k$ . Isto significa que este comportamento não deve ser linear sob a perspectiva do gráfico de *Arrhenius*, e, portanto, o coeficiente  $Q_{10}$  deve ser revisto em casos semelhantes. Esta hipótese foi por nós confirmada ao reavaliar os dados utilizando o modelo proposto  $Q_{d10}$ , os resultados estão expressos na Figura 16A.

Ambos os casos analisados pelos autores apresentaram comportamento do tipo *super-Arrhenius* ( $d > 0$ ). Além disso, o coeficiente  $Q_{d10}$  (Figura 16B) e as energias de ativação (Figura 16C) na faixa de 0 a 20°C mudaram abruptamente com a temperatura. Os biomas de floresta temperada ( $d = 0,2257$ ) foram mais sensíveis



às mudanças de temperatura do que as florestas temperadas decíduas de folha larga ( $d = 0,2173$ ). Em ambos os casos analisados a razão das taxas foi  $1.6 < \frac{k_d(T_2)}{k_d(T_1)} < 2,6$  entre 15 e 40°C, conforme mostrado na Figura 16B. Esses resultados indicam que o parâmetro de deformação ( $d$ ) é um indicador de sensibilidade térmica na respiração das plantas, ou seja, a sensibilidade aumenta com o aumento de  $d$ .

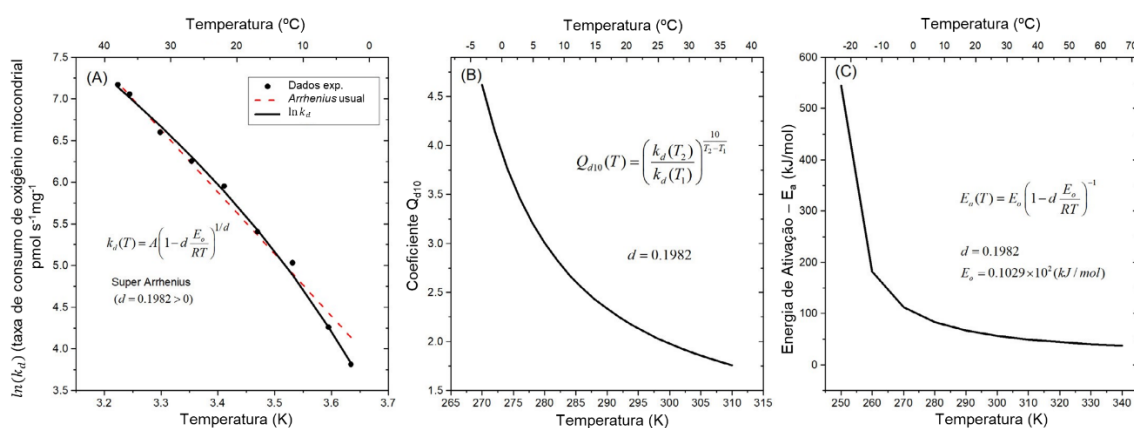


**Figura 16:** Convergência na resposta à temperatura da respiração foliar entre biomassas e tipos funcionais de plantas por meio do modelo de coeficiente deformado ( $Q_{d10}$ ). Gráfico de Arrhenius deformado das taxas de respiração das plantas para biomassas florestais temperadas e tipos funcionais de plantas, incluindo plantas decíduas temperadas de folha larga. (A) Gráfico de Arrhenius deformado, (B) coeficiente  $Q_{d10}$  e (C) energia de ativação deformada.

### 2.3.6 Consumo de Oxigênio nas Mitocôndrias do Fungo *Heteroclito Eurythermal Killifish*

A temperatura ambiente é um fator extremamente importante nos processos envolvidos no metabolismo dos organismos vivos<sup>152,157-161,166,194,195,197,204</sup>. Em particular, a temperatura desempenha um papel importante na distribuição e abundância de uma espécie, afetando o metabolismo aeróbico. Com base nos dados obtidos por Schulte<sup>197</sup> e Fänge e colaboradores<sup>204</sup>, agora demonstramos que a taxa de consumo de oxigênio nas mitocôndrias do fungo *eurythermal killifish* *Fundulus heteroclitus* não pode ser descrita usando a lei de Arrhenius.

A Figura 17A mostra os dados experimentais (pontos sólidos), o gráfico de *Arrhenius* padrão (linha tracejada) e o modelo aqui proposto (linha sólida). A taxa de respiração descrita pelo modelo proposto  $Q_{d10}$  ajusta-se perfeitamente aos dados experimentais com um parâmetro de deformação de  $d = 0,1982$ , o que implica em um comportamento do tipo *super-Arrhenius*. O coeficiente  $Q_{d10}$  indica uma razão de taxa de  $1,5 < \frac{k_d(T_2)}{k_d(T_1)} < 4,5$  entre 0 e 35°C, conforme mostrado na Figura 17B. A energia de ativação mudou abruptamente abaixo de 0°C (Figura 17C). Esses resultados indicaram que a lei de *Arrhenius* não se aplica e não é preditiva na região de temperatura fora daquela estabelecida pelos dados experimentais.



**Figura 17:** Gráfico de Arrhenius deformado para taxa de consumo de oxigênio das mitocôndrias do fungo eurythermal killifish *Fundulus heteroclitus*: (A) dados experimentais, gráfico de Arrhenius padrão (linha tracejada) e o modelo proposto (linha contínua). Os parâmetros de deformação obtidos por ajuste aos dados experimentais são  $d = 0.1982$ ,  $E_o = 0.1029 \times 10^2$  kJ/mol e  $A = 0.3380 \times 10^7$ ; (B) coeficiente  $Q_{d10}$ ; (C) energia de ativação deformada.



## CAPÍTULO 3

### **DEEP LEARNING NA ANÁLISE DE TÉCNICAS ESPECTROSCÓPICAS**

#### **3.1 Introdução**

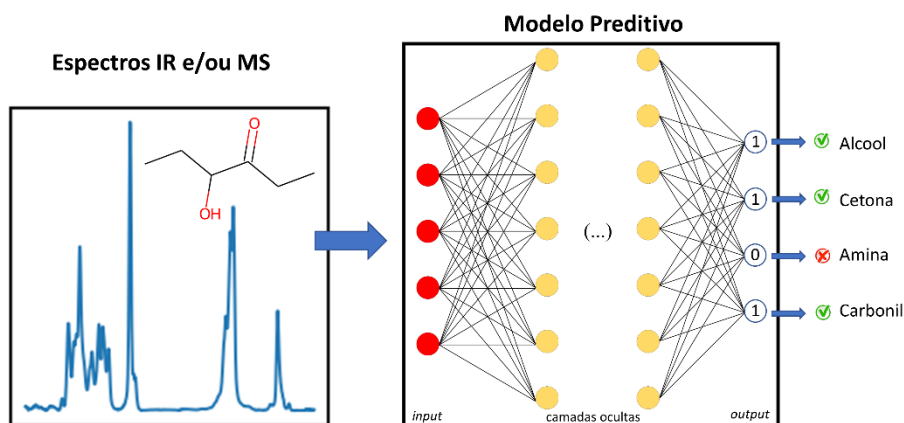
Propriedades físicas e químicas de uma dada substância estão diretamente ligadas à sua forma estrutural (arranjo atômico). Para uma estrutura molecular reconhecida, os químicos podem identificar subestruturas de repetição chamadas grupos funcionais e com isso fazer inferência acerca de diversas destas propriedades. Na química, entretanto, para entidades químicas não conhecidas, o estado da arte na identificação de propriedades estruturais recai sobre a habilidade de espectroscopistas em analisar técnicas espectrográficas, como: espectrometria de massa (MS, do inglês: *Mass Spectrometry*) e infravermelho (IR, do inglês: *Infrared*). Normalmente este é um trabalho intuitivo, no qual um químico identifica manualmente padrões a partir de regras previamente estabelecidas. Pode ser um processo lento e é suscetível a erros, principalmente para moléculas grandes como proteínas ou misturas complexas. Para estruturas conhecidas, existem técnicas computacionais de identificação que sobrepõem e comparam os picos do espectro analisado com um banco de dados<sup>205</sup>, mas a identificação de estruturas desconhecidas ainda é em grande parte um trabalho manual intuitivo dos químicos.

No contexto da inteligência artificial, sabe-se que um dos maiores potenciais recentes desta área envolve a resolução de problemas com soluções intuitivas<sup>2</sup>, em outras palavras, tarefas que são realizadas facilmente por pessoas, mas que são difíceis de serem descritas formalmente, como o caso citado de análises espectroscópicas. Neste sentido, foram aqui desenvolvidos modelos de *Deep Learning* para tratamento e análises automáticas de dados espectrais.

Sabendo que modelos de DL são treinados e aprendem através de um conjunto de dados, para o desenvolvimento deste trabalho realizou-se uma busca automatizada (*web scraping*) no Instituto Nacional de Ciência e Tecnologia dos Estados Unidos (NIST, do inglês: *United States National Institute for Science and Technology*)<sup>206</sup>, onde foram reunidos espectros IR e MS para 1898 compostos. Foram selecionadas moléculas que apresentam picos no espectro IR até  $4000\text{ cm}^{-1}$

e relação carga massa ( $z/m$ ) até 300 no espectro MS, resultando em um total de 1619 compostos. Para explorar ao máximo o banco de dados espectral obtido, definiu-se dois tipos de tratamento e investigação: classificação de características estruturais moleculares (seção 3.3); e reconstrução espectral (seção 3.4). Para a classificação, a ideia é que os espectros funcionem como descritores moleculares e dessa forma os modelos possam realizar inferências acerca de características estruturais das moléculas (ver Figura 18). Na reconstrução espectral a ideia é o inverso: caracterizadores moleculares serão utilizados para reconstruir os espectros.

Para validar os modelos aqui propostos, será apresentada uma série de métricas estatísticas além de comparações de resultados para diversas estruturas de *input* e *output* e arquiteturas de rede. Adicionalmente uma aplicação *web* denominada “*SpectralAI*” foi desenvolvida e os modelos poderão ser testados facilmente por qualquer usuário (mais informações na seção de apêndice B). Espera-se que o desenvolvimento destes modelos e metodologia de análise seja um passo para o uso futuro na detecção analítica autônoma.



**Figura 18:** Modelo preditivo genérico desenvolvido neste trabalho.

Antes de apresentar os modelos desenvolvidos, neste Capítulo 3 será realizada uma abordagem geral sobre o tema inteligência artificial introduzindo as metodologias e bases teóricas utilizadas no desenvolvimento dos modelos.

## 3.2 Bases Teóricas e metodologias

### 3.2.1 Inteligência artificial

Para nos situar no contexto atual da IA, vamos fazer uma breve abordagem histórica. O trabalho de Mcculloch e colaboradores (1943)<sup>207</sup> foi o pioneiro no

desenvolvimento de algoritmos computacionais baseados no funcionamento do cérebro humano, mais especificamente dos neurônios: os autores criaram um modelo computacional chamado “Lógica do Limiar”, do inglês: *threshold logic*. Em 1958 a lógica do limiar é aprimorada com a criação do conceito do *perceptron*<sup>208</sup> um neurônio matemático baseado em uma rede neural computacional de duas camadas. Este modelo abriu caminho para abordagens focadas na aplicação de Redes Neurais Artificiais (RNA) em IA.

Nas décadas seguintes diversos algoritmos baseados em RNA's foram propostos<sup>209,210</sup>, mas a escassez de dados e o alto custo computacional foram cruciais na lenta evolução desta área. Em 1986, um artigo publicado na revista *Nature*<sup>211</sup> revolucionou o treinamento de redes neurais com a técnica “*back-propagation*”. Esta técnica consiste em propagar o erro no sentido contrário do fluxo de dados da rede, descobrindo o quanto de erro cada neurônio da rede é responsável e então atualizar os pesos de modo a diminuir este erro, dando pesos mais altos a neurônios com menores erros.

Esta técnica reacendeu o interesse de pesquisadores pelo tema e já em 1989 tem-se a publicação da primeira “Rede Neural Profunda” (DNN, do inglês: *Deep Neural Network*) que reconhecia dígitos escritos a mão<sup>212</sup>. Em 1992 pesquisadores publicam a *Cresceptron*<sup>213</sup>, uma rede que reconhece objetos 3-D automaticamente. Em 1995 surge as “Máquinas de Vetor de Suporte” (SVM's, do inglês: *Support Vector Machines*)<sup>214</sup>, algoritmos de reconhecimento e mapeamento não supervisionado de dados. Em 1997 publica-se a tradicional “Memória Longa de Curto Prazo” (LSTM, do inglês: *Long short-term memory*)<sup>215</sup> a principal rede neural recorrente que até hoje é a base utilizada em *softwares* de reconhecimento de voz em *smartphones*<sup>216</sup>. Em 1998 Yan LeCun, o criador da técnica *back-propagation*, publica uma nova técnica chamada “Aprendizado Baseado em Gradiente” (GBL, do inglês: *Gradient-based Learning*)<sup>217</sup>, que consiste em um algoritmo estocástico que utiliza o gradiente descendente aliado ao *back-propagation* para agilizar e refinar o treinamento de DNN's.

Em contraste ao rápido desenvolvimento de algoritmos baseados em redes neurais a partir de meados de 1980, a capacidade de processamento dos computadores ainda era muito baixa até por volta do ano 2000: os tempos de treinamento eram contados em dias ou semanas inviabilizando a utilização dos

algoritmos de forma aplicada. Foi a partir do desenvolvimento e melhoramento das placas de vídeo ou “Unidades de Processamento Gráfico” (GPU’s do inglês: *Graphics Processing Unit*) na década seguinte que os algoritmos de DL baseados em redes neurais começam a fazer parte do nosso dia a dia.

Em 2009 foi criada a *ImageNet*<sup>218</sup> um banco de dados com mais de 14 milhões de imagens rotuladas, disponíveis para pesquisadores, professores e estudantes. Em 2012 foi criado um desafio<sup>219</sup> para desenvolvimento de algoritmos de classificação de imagens baseados na *ImageNet*. Vários algoritmos foram desenvolvidos, como: a AlexNet<sup>220</sup>, as VGG’s<sup>221</sup> e as ResNet’s<sup>222</sup>. Uma análise realizada em 2017<sup>223</sup> definiu as ResNet’s como o estado da arte do reconhecimento de imagens devido a sua maior acurácia e densidade de acurácia dentre todas as redes observadas. Nas próximas sessões serão abordados detalhadamente aspectos acerca dos neurônios matemáticos e redes neurais.

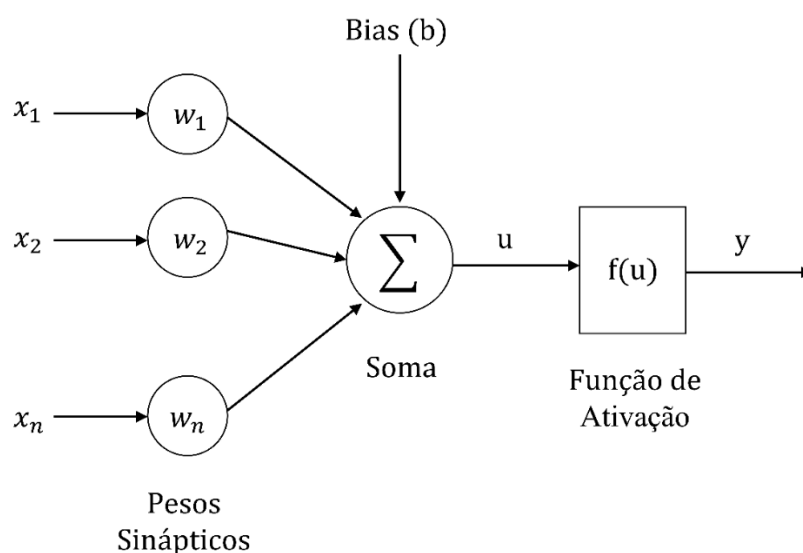
### 3.2.2 O Neurônio Matemático

Um neurônio matemático, artificial ou *perceptron*, é um componente que recebe um ou mais sinais (*input*) e devolve um único sinal de saída (*output*), está representado na Figura 19.

Os sinais  $x_n$  recebidos por um neurônio matemático são análogos à estímulos recebidos por um neurônio biológico. Dentre esses estímulos, alguns irão causar maior ou menor excitação do neurônio receptor, essa medida de excitação é representada pelos pesos sinápticos  $w_n$ . Quanto maior o peso, mais excitatório é o estímulo. Quando os sinais de entrada  $x_n$  chegam nos neurônios, são multiplicados pelos pesos sinápticos  $w_n$  correspondentes, e então é feita uma soma ponderada. Um valor de polarização chamado bias ( $b$ ) é geralmente incluído ao somatório com o intuito de aumentar o grau de liberdade da função e conseqüentemente a capacidade de aproximação (aprendizagem) da rede. Ao final deste processo um sinal de saída ( $u$ ) é gerado. Este sinal de saída é enviado para a função de ativação  $f(u)$  gerando o sinal de saída do neurônio ( $y$ ). O processo total é descrito na Equação (38).

$$y = f(u) = f\left(\sum (w_n * x_n) + b\right) \quad (38)$$

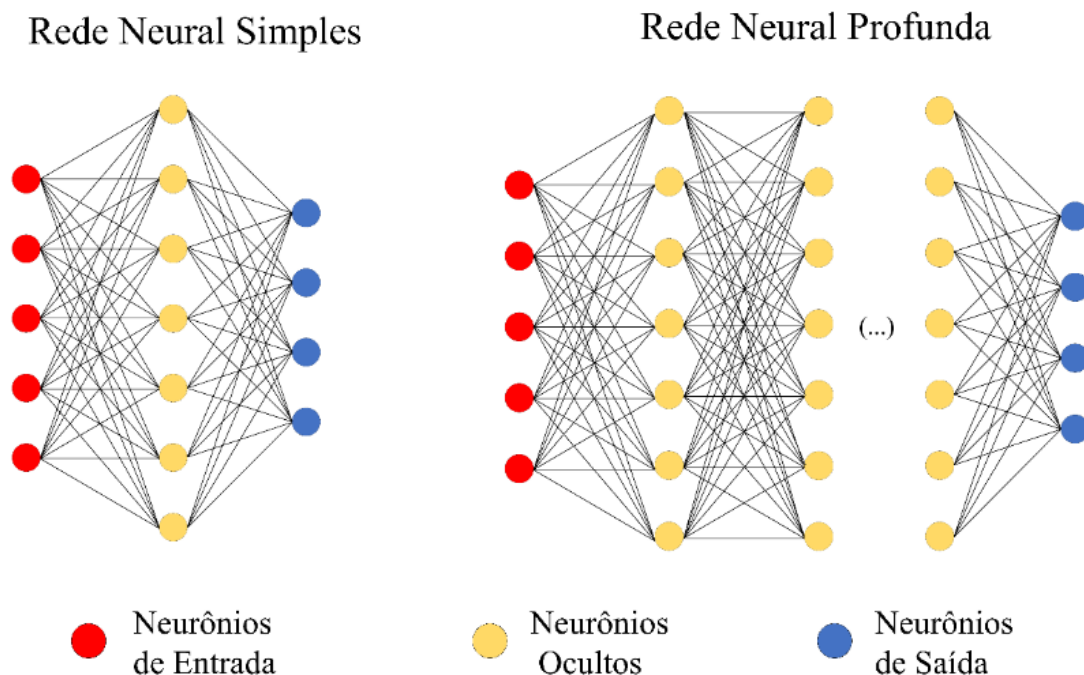
*Perceptrons* quando associados em camadas em sequência dão origem as RNA's e podem descrever diversos tipos de comportamento utilizando um modelo de aproximação universal, análogo a uma soma de polinômios não lineares



**Figura 19:** Representação de um neurônio matemático.  $x_n$  são os sinais de entrada,  $w_n$  são os pesos,  $(u)$  é o resultado da combinação linear,  $f(u)$  é a função de ativação e  $(y)$  é o sinal de saída (*output*).

### 3.2.3 Redes Neurais Artificiais

Redes Neurais Artificiais são algoritmos inspirados no cérebro humano, mais especificamente nos neurônios e suas interconexões. São basicamente algoritmos de ML e DL que apresentam *perceptrons*. RNA's com várias camadas de neurônios (*Multilayer Perceptrons*) são consideradas aproximadores universais<sup>224</sup>. Existem vários tipos de RNA's, aqui será dado foco as Redes Neurais Diretas (do inglês: *feedforward network*) nas quais o fluxo de informação se move em apenas uma direção – para frente – da camada inicial até a camada final. Este fluxo é conhecido como propagação direta. Neste tipo de rede não há *loops* ou ciclos, como acontece nas redes recorrentes (como a LSTM<sup>215</sup>) que necessitam da retroalimentação. Uma RNA direta pode ser definida como um conjunto de neurônios de entrada, um conjunto de neurônios ocultos e um conjunto de neurônios de saída (Figura 20).



**Figura 20:** Representações genéricas de RNA's diretas.

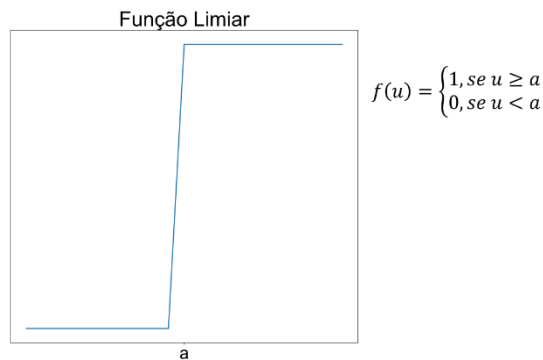
Dentre as diversas tarefas que uma RNA pode desempenhar, daremos destaque para as redes de regressão (que fazem inferência de um valor pretendido de acordo com os dados de entrada) e classificação (que identificam presença ou ausência de características na instância de entrada).

### ***Funções de Ativação***

Para que uma RNA seja um aproximador universal ela deve se adequar aos mais variados tipos de comportamentos. Se os neurônios realizassem apenas combinações lineares  $u = [\sum(w_n * x_n) + b]$  a rede seria uma sequência de combinações lineares, e, portanto, também seria uma combinação linear. É para isto que existem as chamadas funções de ativação. Além de permitir a descrição de comportamentos não lineares, as funções de ativação diminuem o impacto causado por cada peso sináptico na saída final da rede, refinando a excitação dos neurônios e melhorando a capacidade de aproximação ou aprendizagem da RNA.

As funções de ativação são de extrema importância nas RNA's, pois elas basicamente decidem se um neurônio deve ser ativado ou não, ou seja, se a informação fornecida é relevante para o neurônio em questão ou deve ser ignorada. Além disso, são funções diferenciáveis e assim simplificam a determinação do erro pelo *back-propagation* ao possibilitarem a determinação de seu gradiente para cada peso, contribuindo para o refino do treinamento da rede.

A função de ativação mais simples é a função limiar (discutida no início da seção 3.2.1), seu comportamento é descrito na Figura 21. Para esta função, o valor da saída  $y$  do neurônio será igual à 1 (ou ativado) se o valor de entrada  $u$  for igual ou maior que uma dada constante  $a$ , e será igual à 0 (ou não ativado) caso a entrada seja menor que esta mesma constante. A função limiar basicamente decide se um sinal é relevante ou deve ser ignorado, ou seja, se aquele neurônio foi ativado, ou não.



**Figura 21:** Função Limiar

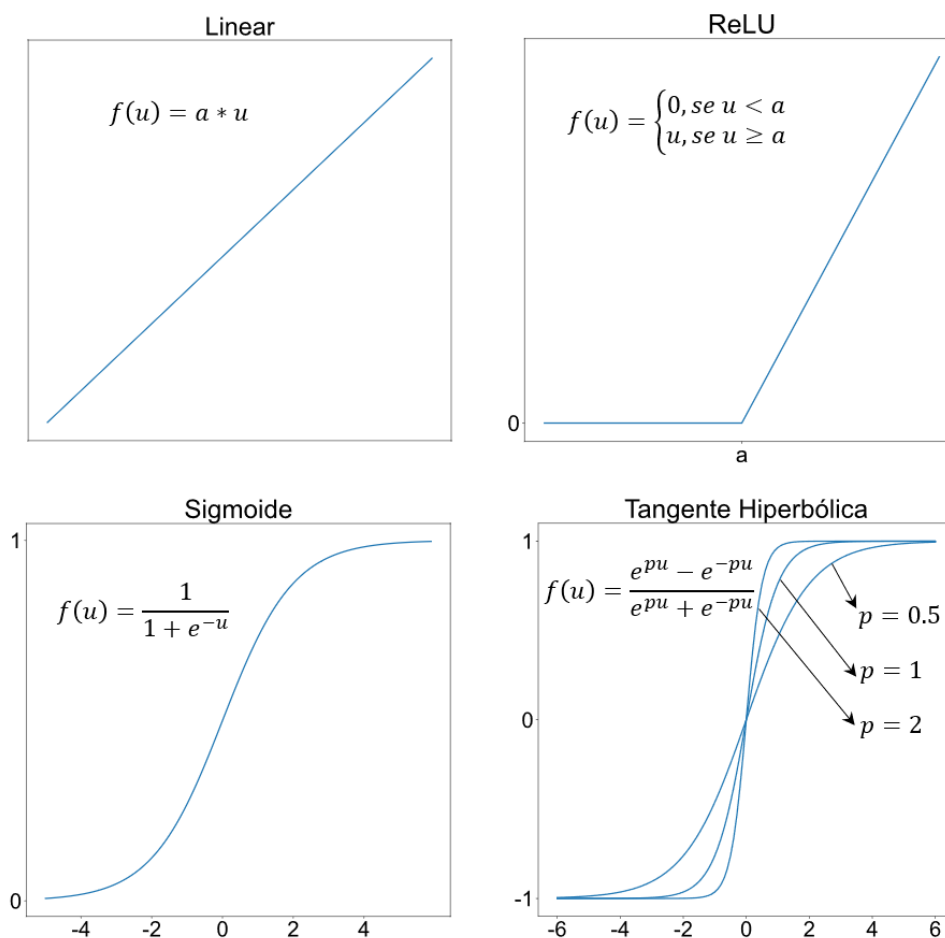
Atualmente as funções de ativação mais utilizadas em RNA's são aquelas que retornam uma probabilidade de ativação: ao invés de retornar os valores 0 ou 1, retornam valores reais dentro de um intervalo, representando a relevância do sinal de entrada  $u$  ou a probabilidade de ativação. Dentre as mais comuns podemos citar: a função linear; função ReLU, função sigmóide; e a função tangente hiperbólica, descritas nas Equações (39), (40), (41) e (42), respectivamente. Na Equação (42),  $p$  é um parâmetro arbitrário. Os comportamentos destas funções estão expressos na Figura 22.

$$f_{lin}(u) = au \quad (39)$$

$$f_{ReLU}(u) = \begin{cases} 0, & \text{se } u < a \\ u, & \text{se } u \geq a \end{cases} \quad (40)$$

$$f_{sig}(u) = \frac{1}{1 + e^{-u}} \quad (41)$$

$$f_{tanh}(u) = \frac{e^{pu} - e^{-pu}}{e^{pu} + e^{-pu}} \quad (42)$$



**Figura 22:** Comportamento das funções de ativação: linear, ReLU, sigmoide e tangente hiperbólica

### **Treinamento e Hiperparametrização**

Um passo importante no desenvolvimento de RNA's é o treinamento, que consiste em modelar o comportamento do sinal de saída de uma rede para que possa gerar ou classificar informações. Para isto é preciso calibrar os pesos sinápticos  $w_n$  (de forma similar à busca de parâmetros da função custo realizado pelo GSA).

Neste processo os pesos são inicialmente aleatorizados (ou herdados de uma outra rede já treinada para outra tarefa). A cada "lote" de dados que flui sobre a rede, os pesos  $w_n$  e o bias  $b$ , Equação (38), são atualizados sempre visando minimizar o erro, este processo de atualização dos pesos é realizado por otimizadores. O GSA é um otimizador, pois apresenta o seu próprio método de otimização. Outros otimizadores comuns em ML são o Gradiente Descendente Estocástico (SGD, do inglês: *Stochastic Gradient Descent*) e o Adam<sup>225</sup>.



A função  $\chi^2$  Equação (32) é uma função erro bastante comum usada em problemas de regressão. Para problemas de classificação comumente se usa a função entropia cruzada<sup>226</sup>.

Ao treinar uma rede, escolher corretamente os otimizadores e a função erro são de grande importância, mas sozinhos não garante sucesso do treinamento. Existem uma série de hiper parâmetros que devem ser ajustados ao problema em questão, como: a taxa de aprendizagem (do inglês *learning rate* é a grandeza com que os pesos serão atualizados: se for muito baixa serão necessárias demasiadas etapas de treinamento; se a for muito alta os pesos irão se alterar muito bruscamente); o número de camadas do modelo; a quantidade de neurônios em cada camada; e o número de ciclos de treinamento. Quanto mais complexo o problema, mais características os dados apresentam e maior número de camadas e neurônios serão necessários.

### **Classificadores Multi Classe e Multi Rótulo**

Um algoritmo de classificação multi rótulo (do inglês: *Multilabel Classification*) é capaz de determinar à quais classes uma instância pertence, diferente de algoritmos de classificação binária multi classe, onde o objetivo é determinar à qual classe (única) aquela instância pertence. Nestes classificadores geralmente o sinal de saída da rede é um vetor de probabilidades (valores entre 0 e 1). Cada posição do vetor representa uma classe, enquanto o valor é a probabilidade de a classe estar presente na instância (ver exemplo de aplicação na seção 3.3.2). Em classificação binária multi classe a classe com maior probabilidade será o resultado da predição; em classificação multi rótulo é necessário um critério para definir quais classes estão presentes na instância, como por exemplo a determinação de limiares.

### **Métricas de validação**

Para avaliar a eficiência de um modelo de classificação multi rótulo utiliza-se funções estatísticas como: acurácia [*Acc*, Equação (43)], *recall* [*Rec*, Equação (44)], precisão [*Prec*, Equação (45)], especificidade [*Spec*, Equação (46)] e pontuação F1 [*F1 – Score*, Equação (47)]<sup>227</sup>. Cada uma dessas métricas avalia uma característica diferente da rede, iremos discutir todas elas nesta seção. Quando um modelo prediz corretamente uma classe contida na instância entende-se que ocorreu um Positivo Verdadeiro (*TP*, do inglês: *True Positive*) enquanto a predição de uma classe que

não está contida na instância representa um Falso Positivo (*FP*, do inglês: *False Positive*). De forma análoga, quando uma classe presente na instância não é predita pelo modelo, tem-se um Falso Negativo (*FN*, do inglês: *False Negative*), enquanto a não detecção de uma classe que de fato não existe na instância representa um Negativo Verdadeiro (*TN*, do inglês: *True Negative*). Munidos destes conceitos iremos definir as métricas citadas anteriormente<sup>227</sup>.

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \quad (43)$$

$$Rec = \frac{TP}{TP + FN} \quad (44)$$

$$Prec = \frac{TP}{TP + FP} \quad (45)$$

$$Spec = \frac{TN}{TN + FP} \quad (46)$$

$$F1 - Score = \frac{2 * Prec * Rec}{Prec + Rec} \quad (47)$$

Essas métricas são calculadas item a item, classe por classe: a média das classes representa a métrica total.

A acurácia, Equação (43), é a proporção das predições verdadeiras em relação ao total de predições realizadas. Geralmente não é suficiente para avaliar a eficiência de um classificador multi rótulo, por exemplo: considere um conjunto de dados em que 10% dos itens apresentem uma certa classe, se utilizarmos uma rede enviesada que não prevê esta classe nunca, ainda sim esta rede teria uma acurácia de 90%.

O *recall*, Equação (44), representa a proporção de acertos de previsão realizadas (*TP*) em todas as instâncias que apresentam uma certa classe (*TP + FN*); enquanto a precisão, Equação (45), é a proporção de predições corretas (*TP*) dentre todas as vezes que uma classe foi predita (*TP + FP*): considere um conjunto de dados com 200 itens, onde 100 itens apresentem uma classe específica; e considere um algoritmo que detecta corretamente esta classe em 80 dos itens que realmente apresentam esta classe; ao mesmo tempo, o algoritmo detectou esta classe em outras 100 instâncias que originalmente não possuem a classe em questão. Neste

caso, temos:  $TP = 80$ ,  $FP = 100$ ,  $TN = 20$  e  $FN = 0$ . O *recall* neste caso será 80%, enquanto a precisão será 44%. A especificidade, Equação (46), é análoga a precisão, mas avalia a proporção com que a rede identifica corretamente a ausência de certa classe, seu valor neste caso é de 16,66%.

A importância de cada métrica depende do modelo e do problema abordado, por exemplo: qual o mais perigoso para uma rede treinada para identificação de câncer em pacientes: identificar um câncer não existente (*FP*), ou deixar de identificar um câncer existente (*FN*)? A resposta correta é provavelmente a segunda e neste caso devemos garantir uma boa medida para o *recall* desta rede; mas se a resposta correta fosse a primeira, deveríamos garantir que nossa rede tenha uma boa precisão e especificidade. Além disso, no geral quanto melhor o *recall*, precisão e especificidade, melhor tende a ser a acurácia.

A última e mais importante métrica dos classificadores é a *F1-Score*: uma relação entre *recall* a precisão. Seu valor máximo é 1 ou 100%, indicando *recall* e precisões perfeitos. Em suma, bons valores de *F1-Score* indicam que uma rede está falhando pouco em todos os quesitos. O *F1-Score* é geralmente utilizado para comparação de desempenhos entre redes. Entretanto, é necessário pontuar que apenas essa métrica não é suficiente para indicar a eficiência de uma rede, sendo necessário análises de várias métricas em um contexto geral para se chegar a esta conclusão.

### 3.2.4 Redes Neurais Convolucionais

Uma Rede Neural Convolucional (CNN, do inglês: *Convolutional Neural Network*) é um RNA profunda que utiliza dados matriciais (que podem apresentar diferentes tamanhos) como *input*, atribuindo importância a aspectos destas matrizes, sendo, portanto, capaz de diferenciá-los. Imagens são matrizes de pixels que podem apresentar 1 ou mais canais de cores (3 canais no caso RGB, por exemplo) e foram o foco principal no desenvolvimento deste tipo de rede. As CNN's se diferem das RNA's convencionais devido a suas camadas iniciais de pré-processamento da imagem, que precedem as camadas de neurônios. Em métodos primitivos de IA o pré-processamento das imagens era algo complexo, pois os filtros utilizados para extrair características das imagens eram feitos a mão para cada conjunto de

imagens de entrada. Já as CNN's, com treinamento suficiente, são capazes de aprender esses filtros (características).

Dentre as operações de pré-processamento da imagem temos: o *pooling* (operação responsável por diminuir a dimensionalidade da imagem, camadas de pooling são essenciais para otimizar o tempo de treinamento da rede); achatamento ou *flating* (geralmente realizada pela camada final de pré-processamento da imagem onde as matrizes da imagem são “achatadas” e transformadas em vetores unitários que são enviados como *input* para primeira camada de neurônios); e as convoluções. Esta última é a operação mais importante, responsável pela extração de características na imagem, e será abordada com mais detalhes.

A operação de convolução é a aplicação de um filtro em uma imagem, sendo, portanto, responsável por extrair as características dessa imagem dando à rede a capacidade de diferenciar e classificar essas características. Matematicamente a operação de convolução é uma sequência de produtos escalares de matrizes, na qual a matriz do filtro utilizado (de tamanhos variados 3x3, 4x4 etc.) varre a matriz de pixels da imagem. A média do produto escalar para cada superposição entre o filtro e uma região da imagem da origem a um novo pixel da imagem gerada. Desta forma, cada filtro utilizado gera uma nova imagem, diferentes entre si, e cada uma dessas imagens resultantes carrega consigo uma característica da imagem inicial.

Diferentes combinações entre camadas de pré-processamento e de neurônios dão origem a diversas arquiteturas avançadas de CNN's com diferentes capacidades de aprendizagem. Atualmente as CNN's são o estado da arte dos algoritmos de DL para classificação de imagens. As redes citadas na seção 3.2.1: VGG, AlexNet e ResNet são exemplos de CNN's.

### **3.2.5 Codificação de moléculas para uso em ML e DL**

Atualmente boa parte das pesquisas em química se dedicam a projetar novas moléculas com propriedades específicas, e, embora haja diversos trabalhos que visam novas estratégias de *design*, ainda é muito comum a utilização de métodos aleatórios para desenvolver moléculas de interesse. Um dos maiores objetivos da utilização de ML e DL na química é substituir estes métodos aleatórios por uma busca direcionada, no qual modelos de predição possam ser utilizados para prever

propriedades moleculares e assim aumentar a eficiência no desenvolvimento de novos materiais e técnicas de análise.

Para isso, o primeiro passo é construir métodos que transformem moléculas em vetores numéricos que possam ser passados para os algoritmos de ML e DL, chamados “caracterizadores moleculares” (do inglês, *molecular featurizations*). Moléculas são entidades complexas, por isso existem diversas técnicas para caracterizá-las, incluindo vetores de descritores químicos, representações de gráficos 2D, representações 3D, representações de funções de base orbital e muito mais. Uma vez caracterizada, uma molécula pode ser “aprendida” por um modelo. Dois caracterizadores foram utilizados como descritores moleculares para gerar espectros moleculares de IR e MS: impressões digitais de conectividade estendida (ECFPs, do inglês: *Extended-connectivity fingerprints*) e matriz de *coulomb*.

### **ECFPs**

Impressões digitais químicas (do inglês, *molecular fingerprints*) são vetores contendo valores de 0 e 1 que representam a presença ou ausência de características específicas em uma molécula. Foram inicialmente desenvolvidas para facilitar a busca de estruturas em bancos de dados e atualmente são utilizadas em tarefas de ML para busca de similaridade e clusterização<sup>228</sup>.

ECFPs é uma metodologia mais recente<sup>229</sup> que visa descrever características moleculares relevantes para atividade química e é utilizada bastante em modelagem molecular. Podem ser calculadas rapidamente e transformam uma molécula de tamanho arbitrário em vetores de tamanho fixo, facilitando a comparação de similaridade entre moléculas uma vez que só é preciso comparar os elementos correspondentes dos vetores. Cada elemento do vetor representa a presença ou ausência de uma característica específica definida por um arranjo local de átomo, considerando propriedades atômicas como: elemento atômico e número de ligações covalentes.

### **Matriz de Coulomb**

Matriz de *Coulomb* é um caracterizador simples que descreve as interações eletrostáticas entre os átomos de uma molécula ao codificar cargas nucleares e coordenadas cartesianas correspondentes em uma matriz. A regra de formação para a matriz de *Coulomb* obedece a Equação (48):

$$M_{ij}^{Coulomb} = \begin{cases} 0.5Z_i^{2.4} & \text{for } i = j \\ \frac{Z_i Z_j}{R_{ij}} & \text{for } i \neq j \end{cases} \quad (48)$$

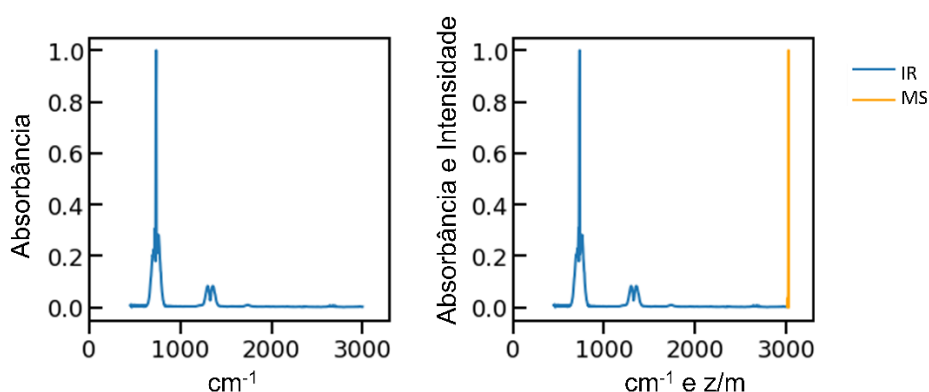
na qual  $i$  e  $j$  são índices atômicos;  $Z_i$  e  $Z_j$  são cargas nucleares e  $R_{ij}$  é a distância entre os átomos  $i$  e  $j$ . Os elementos da diagonal principal podem ser interpretados como a interação de cada átomo com si mesmo, enquanto os elementos fora da diagonal principal representam a repulsão nuclear entre os átomos  $i$  e  $j$ . Neste trabalho a maior molécula do conjunto de dados trabalhado apresenta 72 átomos (sem considerar os átomos de hidrogênio), portanto foi definido o tamanho padrão 72x72 da matriz de *coulomb* para todas as moléculas ao incluir linhas e colunas de zeros para moléculas menores.

### 3.3 Caracterização de grupos funcionais moleculares

Geralmente o resultado de técnicas espectrográficas são expressos em imagens, por isso dois modelos de DL foram desenvolvidos: um que usa dados espectrais brutos (dados tabulares) e outro que utiliza imagens dos espectros (detalhes acerca destes modelos serão apresentados na próxima seção). Ambos os modelos foram treinados a partir dos seguintes conjuntos de dados espectrais:

- i. Espectros MS no intervalo de 0 a 300 ( $z/m$ )
- ii. Espectros IR no intervalo de 0 a 4000  $cm^{-1}$
- iii. Espectros IR no intervalo de 0 a 4000  $cm^{-1}$  + espectros MS no intervalo de 0 e 300 ( $z/m$ ).

No caso (iii) os espectros IR e MS foram concatenados, (ver Figura 23). Os compostos reunidos foram rotulados indicando a presença ou ausência de grupos funcionais e outras características estruturais, através de funções de checagem de fragmentos da ferramenta RDKit<sup>230</sup>. Ambos os modelos foram treinados para predição das características listadas na Tabela 5, que também mostra a quantidade de compostos que possui a determinada característica. Os índices de cada característica, expressos na Tabela 5, serão utilizados como referência nas Figuras e Tabelas ao decorrer deste Capítulo 3.



**Figura 23:** A esquerda um espectro IR e a direita o mesmo espectro IR concatenado com o espectro MS.

**Tabela 5:** Ao meio a lista das classes analisadas, dentre elas: grupos funcionais e fragmentos diversos. A direita a quantidade de compostos que apresentam o fragmento de referência e a esquerda o índice de cada classe.

Índice	Classes	Quantidade de Compostos
1	Álcool	252
2	Ácido Carboxílico Aromático	40
3	Ácido Carboxílico	104
4	Carbonil	698
5	Amina	331
6	Aldeido	108
7	Haleto de alquila	111
8	Anilina	106
9	Anel Benzênico	824
10	Ester	252
11	Heteroátomo de oxigênio (incluindo fenoxi)	564
12	Halogênio	311
13	Cetona	194
14	Metoxi	207
15	Nitro	85
16	Alvos de Para-Hidroxilação	162
17	Fenol	218
18	Alcanos não ramificados de pelo menos 4 membros (exclui alcanos halogenados)	263

### 3.3.1 Treinamento e hiperparametrização

Todos os modelos desenvolvidos foram treinados utilizando a biblioteca *fastai*<sup>231</sup>. Para o modelo que utiliza imagens é necessário utilizar uma CNN: testes

foram realizados crescendo o número de camadas de ResNet's<sup>222</sup> e identificou-se uma melhora das métricas até 34 camadas, sendo este o critério para definição da ResNet34 como a rede utilizada neste trabalho. Para o modelo que utiliza os dados espectrais brutos, o chamado modelo tabular, testes foram realizados crescendo a quantidade de camadas ocultas e o número de neurônios em cada camada, chegando ao ponto ótimo na arquitetura com 5 camadas ocultas com respectivamente 1024, 512, 256, 128 e 64 neurônios nestas camadas. Apenas a rede tabular que utiliza dados de MS apresenta apenas 3 camadas ocultas: 256, 128 e 64, uma vez que o tamanho do seu *input* é 300.

A CNN ResNet34 é um modelo que utiliza imagens como *input* e por isso apresenta uma grande quantidade de parâmetros ajustáveis, somando mais de 22 milhões de parâmetros no total, enquanto a quantidade de parâmetros ajustáveis no modelo tabular adotado é próximo de 5 milhões (o modelo tabular que utiliza dados de MS e tem apenas 3 camadas apresenta cerca de 120 mil parâmetros ajustáveis). Em todos os procedimentos realizados os modelos foram treinados com 80% dos dados, num total de 1294 compostos, enquanto os outros 20% foram separados para testes de validação. O conjunto de testes de validação não é processado pela rede durante o treinamento e não influencia o ajuste dos pesos: é de suma importância pois garante que a rede será testada e validada com compostos que nunca viu antes, prevenindo o *overfitting*.

Os modelos foram treinados em computador pessoal com as configurações: processador i5-8400 2.81 GHZ, 16GB de memória RAM, GPU Nvidia GTX-1060 6GB. O treinamento de cada modelo ocorreu por cerca de 100 passos com a taxa de aprendizagem sendo reajustada a cada 30 passos. O otimizador utilizado foi o Adam<sup>225</sup> e a função erro a Entropia Cruzada<sup>226</sup>.

### 3.3.2 Validação

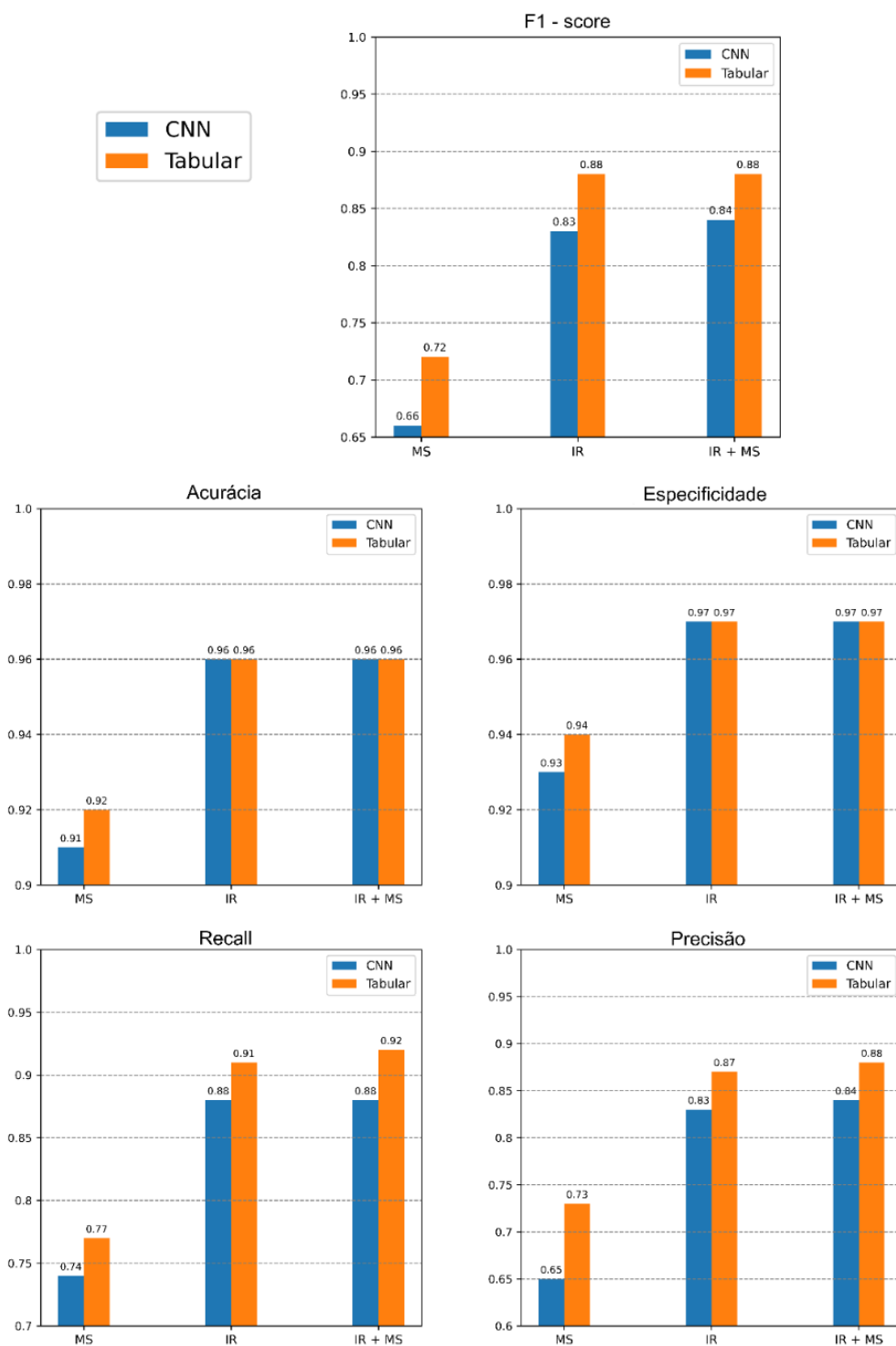
As métricas descritas nas Equações (43) a (47) foram calculadas para os modelos desenvolvidos (CNN e tabular) para os três conjuntos de dados estudados (MS, IR e IR+MS). As médias ponderadas classe a classe estão expressas na Figura 24. Nos dois modelos utilizados (CNN e tabular), o que utiliza dados de MS se mostrou pouco satisfatório na tarefa de classificação, apresentando métricas bem abaixo dos demais. Para os modelos que usam dados de IR e IR+MS as métricas



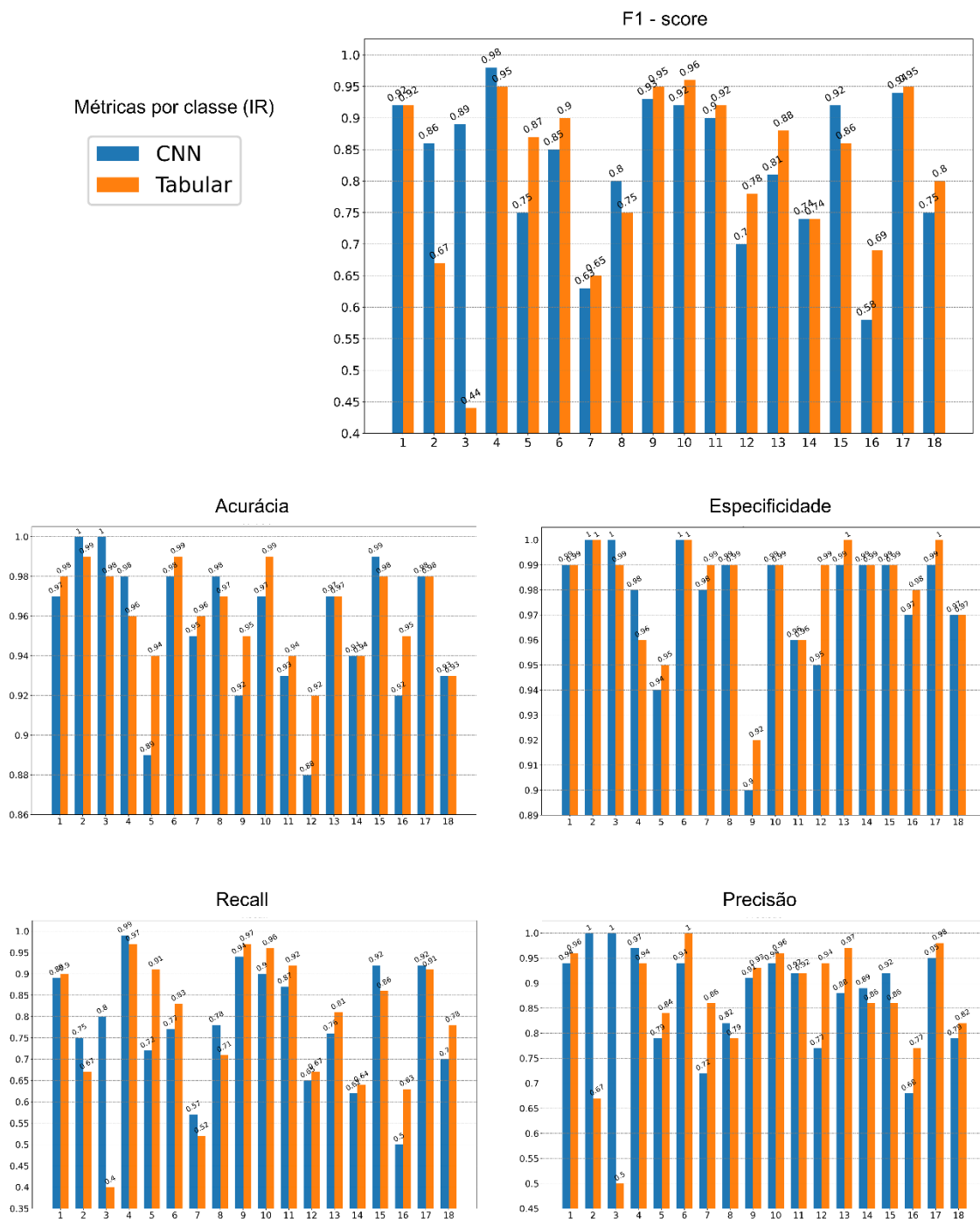
foram bem semelhantes, sendo ligeiramente melhores para o modelo tabular. Entretanto, a diferença entre as métricas ainda é muito baixa (entre 0% e 5%) para afirmar qual modelo é mais eficiente. Uma alternativa é computar as métricas para cada classe individualmente e verificar qual modelo se sai melhor em um maior número de classes. As métricas individuais para cada classe estão expressas na Figura 25 para o conjunto de dados IR e na Figura 26 para IR+MS (não serão apresentadas as métricas individuais para os modelos que utilizam o conjunto de dados MS pois as métricas não foram satisfatórias).

Como discutido na seção 3.2.3, para efeito de comparação entre modelos a métrica *F1-Score* é a melhor escolha. Analisando a Figura 25 para dados de IR, o modelo CNN obteve um melhor *F1-Score* para 5 classes, enquanto o modelo tabular foi melhor na predição de 11 das classes preditas, os dois modelos empataram para outras 2 classes. Na Figura 26 para dados IR+MS, o modelo CNN obteve um melhor *F1-Score* para 6 das classes preditas, enquanto o modelo tabular foi melhor para outras 9 classes e houve um empate para as outras 3 classes.

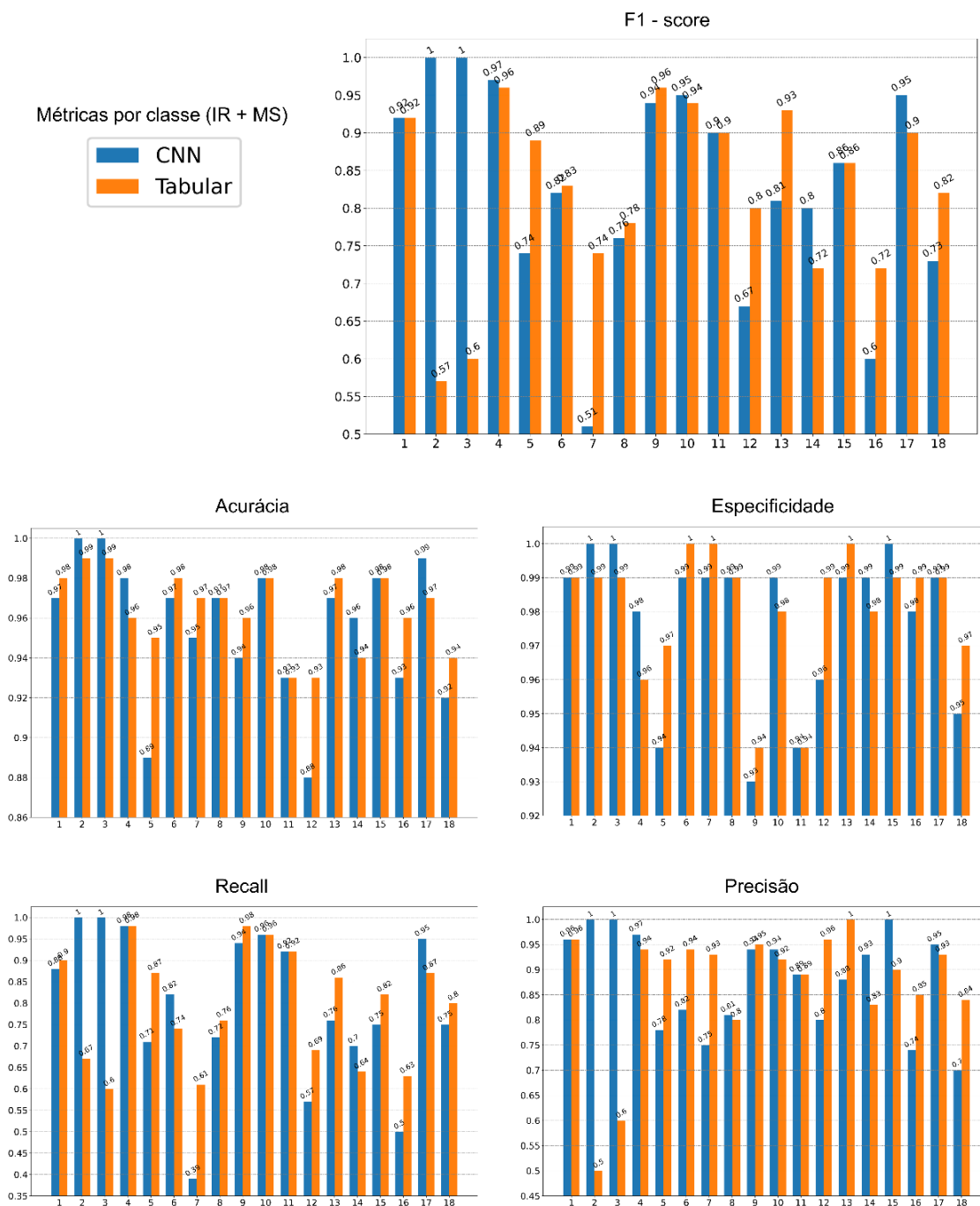
Tanto nas métricas médias (Figura 24) quanto na análise individual de métricas por classe (Figura 25 e Figura 26), o modelo tabular se saiu melhor do que o modelo CNN que utiliza imagens, indicando que este modelo pode ser a melhor escolha para uma predição mais eficiente. Entretanto, dada a proximidade de desempenho entre os modelos não é possível definir um modelo que seja indubitavelmente melhor ou mais apropriado para predição das características estruturais moleculares estudadas. Para ilustrar as semelhanças e diferenças entre as classificações feitas pelos modelos, a Tabela 6 e Tabela 7 destacam quatro exemplos contidos no nosso conjunto de dados de validação e mostra as métricas calculadas para as predições do modelo.



**Figura 24:** Métricas de validação para os modelos CNN e tabular utilizando os diferentes conjuntos de dados espectrais: MS, IR e IR + MS.

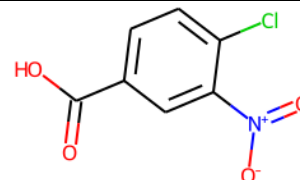
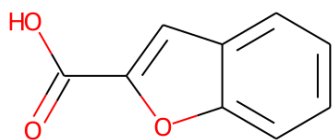


**Figura 25:** Acurácia, recall, especificidade, precisão e F1 – Score para cada classe individualmente, para os modelos CNN e Tabular utilizando o conjunto de dados espectrais de IR.



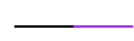
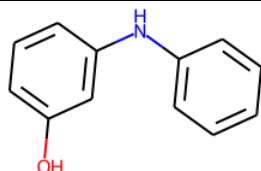
**Figura 26:** Acurácia, recall, especificidade, precisão e F1 – Score para cada classe individualmente, para os modelos CNN e Tabular utilizando o conjunto de dados espectrais de IR + MS.

**Tabela 6:** As moléculas “ácido cumarílico” e “ácido 3-nitro-4-clorobenzóico” fazem parte do nosso dataset de validação e foram avaliadas para os modelos CNN e Tabular treinados com os conjuntos de dados IR e IR + MS. As colunas “Rótulo” são vetores que indicam quais grupos funcionais a molécula apresenta (padrão ouro), enquanto as colunas “Predição” são os vetores de saída dos modelos que indicam quais grupos funcionais foram preditos para aquela molécula. Na parte de baixo da tabela estão expressas as métricas: F1-Score, acurácia, recall, precisão e especificidade calculadas diretamente entre dos vetores de predição e rótulo correspondentes.



Classes	Rótulo	Predição				Rótulo	Predição			
		CNN		Tabular			CNN		Tabular	
		IR	IR + MS	IR	IR + MS		IR	IR + MS	IR	IR + MS
1	0	0	0	0	0	0	0	0	0	
2	1	1	1	1	1	1	1	1	1	
3	1	1	1	1	1	1	1	1	1	
4	1	1	1	1	1	1	1	1	1	
5	0	0	0	0	0	1	1	1	1	
6	0	0	0	0	0	0	0	0	0	
7	0	0	0	0	0	0	0	0	0	
8	0	0	0	0	0	0	0	0	0	
9	1	1	1	1	1	1	1	1	1	
10	0	0	0	0	0	0	0	0	0	
11	0	0	0	0	0	0	0	0	0	
12	0	1	0	0	0	1	1	0	0	
13	0	0	0	0	0	0	0	0	0	
14	0	0	0	0	0	0	0	0	0	
15	0	0	0	0	0	1	1	1	1	
16	1	0	0	1	1	0	0	0	0	
17	0	1	0	0	0	0	0	0	0	
18	0	0	0	0	0	0	0	0	0	
<b>F1 – Score</b>		0,73	0,89	1,0	1,0		1,0	1,0	0,92	0,92
<b>Acurácia</b>		0,83	0,94	1,0	1,0		1,0	1,0	0,94	0,94
<b>Recall</b>		0,80	0,80	1,0	1,0		1,0	1,0	0,86	0,86
<b>Precisão</b>		0,67	1,0	1,0	1,0		1,0	1,0	1,0	1,0
<b>Especificidade</b>		0,85	1,0	1,0	1,0		1,0	1,0	1,0	1,0

**Tabela 7:** As moléculas “m-hidroxidifenilamina” e “iodometano” fazem parte do nosso dataset de validação e foram avaliadas para os modelos CNN e Tabular treinados com os conjuntos de dados IR e IR + MS. As colunas “Rótulo” são vetores que indicam quais grupos funcionais a molécula apresenta (padrão ouro), enquanto as colunas “Predição” são os vetores de saída dos modelos que indicam quais grupos funcionais foram preditos para aquela molécula. Na parte de baixo da tabela estão expressas as métricas: F1-Score, acurácia, recall, precisão e especificidade calculadas diretamente entre dos vetores de predição e rótulo correspondentes.



Classes	Rótulo	Predição				Rótulo	Predição			
		CNN		Tabular			CNN		Tabular	
		IR	IR + MS	IR	IR + MS		IR	IR + MS	IR	IR + MS
1	0	0	0	0	0	0	0	0	0	
2	0	0	0	0	0	0	0	0	0	
3	0	0	0	0	0	0	0	0	0	
4	0	0	0	0	0	0	0	0	0	
5	1	1	1	1	1	0	1	0	0	
6	0	0	0	0	0	0	0	0	0	
7	0	0	0	0	0	1	0	1	1	
8	1	1	1	1	1	0	0	0	0	
9	1	1	1	1	1	0	1	0	0	
10	0	0	0	0	0	0	0	0	0	
11	0	0	0	0	0	1	0	0	0	
12	0	0	0	0	0	1	0	1	1	
13	0	0	0	0	0	0	0	0	0	
14	0	0	0	0	0	0	0	0	0	
15	0	0	0	0	0	0	0	0	0	
16	1	0	0	0	0	0	0	0	0	
17	1	0	1	1	0	0	0	0	0	
18	0	0	0	0	0	0	0	0	0	
<b>F1 – Score</b>		0,75	0,89	0,89	0,75		0,0	0,0	1,0	1,0
<b>Acurácia</b>		0,89	0,94	0,94	0,89		0,83	0,78	1,0	1,0
<b>Recall</b>		0,6	0,8	0,8	0,6		0,0	0,0	1,0	1,0
<b>Precisão</b>		1,0	1,0	1,0	1,0		0,0	0,0	1,0	1,0
<b>Especificidade</b>		1,0	1,0	1,0	1,0		0,94	0,88	1,0	1,0

### 3.4 Espectros moleculares gerados a partir de descritores moleculares

Na seção anterior, o banco de dados espectral reunido foi utilizado para classificar propriedades moleculares, ou seja, os espectros eram *inputs* dos modelos que classificavam informações de interesse, obtidas como *output* desses modelos. Nesta seção iremos utilizar os descritores moleculares abordados na seção 3.2.5 – circular *fingerprint* e matriz de *coulomb* – como *input* para nossos modelos aprenderem a gerar os espectros correspondentes pelo método de regressão, utilizando a função custo erro quadrático médio, expressa na Equação (49):

$$\sum_{i=1}^n [I_{true_i} - I_{pred_i}]^2 / n \quad (49)$$

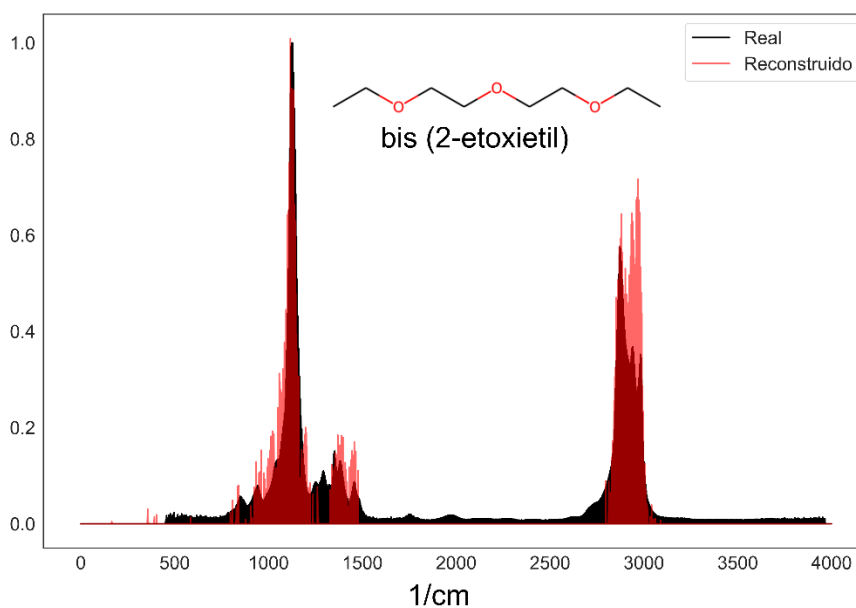
na qual  $I_{true}$  é a intensidade real;  $I_{pred}$  é a intensidade predita;  $i$  é o valor em  $cm^{-1}$  correspondente para o espectro IR e  $z/m$  para MS; e  $n$  é o tamanho do vetor de intensidades (4000 para IR e 300 pra MS).

Diversas arquiteturas foram desenvolvidas e os treinamentos mostraram que os modelos de fato aprendiam até por volta do décimo passo de treinamento, ao apresentar valores cada vez menores da função custo para o conjunto de validação, a partir daí esse valor se manteve constante. Entretanto nenhum dos modelos treinados gerou espectros com qualidade satisfatória. Uma vez que dados espectrais possuem padrões altamente complexos, o tamanho do conjunto de dados utilizado foi um fator limitante no aprendizado destes modelos. O espectro de IR, por exemplo, apresenta 4000 valores de intensidades, enquanto o nosso conjunto de dados contempla pouco mais de 1800 compostos. Provavelmente, com um conjunto de dados maior, os modelos irão desempenhar melhor a tarefa de gerar espectros moleculares.

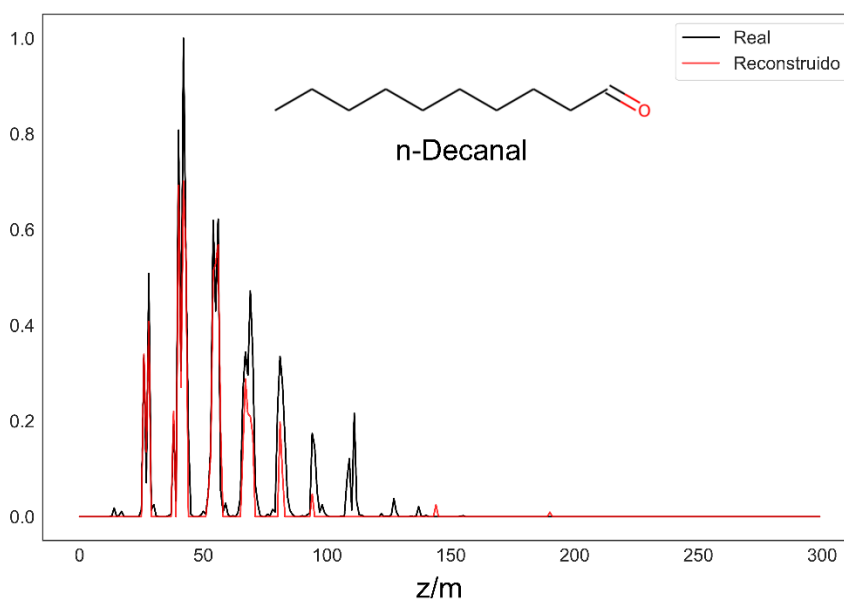
O modelo que obteve o melhor aprendizado foi o modelo que utilizou o descritor “circular *fingerprint*”, obtendo um erro quadrático médio de 0.005 para reconstrução do espectro MS e 0.007 para a reconstrução do espectro IR. Este descritor é um vetor unitário de tamanho 1024, portanto foi desenvolvido um modelo *autoencoder* tabular no qual, para reconstruir o espectro de IR (de tamanho 4000), o *encoder* apresenta 2 camadas ocultas com 512 e 256 neurônios e o *decoder* 3 camadas ocultas com 256, 512 e 2048 neurônios respectivamente. Para reconstruir o espectro de massa (MS, de tamanho 300) o *autoencoder* foi desenvolvido com o *encoder* contendo 2 camadas ocultas com 256 e 128 neurônios e o *decoder* com 2

camadas ocultas de 128 e 256 neurônios respectivamente. O modelo que reconstruí espectros IR tem mais de 10 milhões parâmetros ajustáveis enquanto o modelo que reconstrói o espectro MS tem mais de 800 mil.

A Figura 27 e Figura 28 mostram, respectivamente, as melhores reconstruções para os espectros IR e MS. É possível observar que os modelos obtiveram algum aprendizado, entretanto não foi suficiente para gerar espectros de qualidade.



**Figura 27:** Espectro IR da molécula bis (2-etoexietil) obtido no NIST e a reconstrução gerada pelo modelo.



**Figura 28:** Espectro MS da molécula n-Decanal obtido no NIST e a reconstrução gerada pelo modelo.



## CONSIDERAÇÕES FINAIS E PERSPECTIVAS

A baixas temperaturas fatores energéticos passam a ter uma menor contribuição na dinâmica de um processo e, em alguns casos, outros regimes passam a ter maior influência modificando sensivelmente a velocidade destes processos, provocando desvios no gráfico de *Arrhenius*. No intuito de observar estas mudanças de regime e descrever a influência destes fenômenos na cinética de um processo, foi aqui proposto a função transitividade  $\gamma(T)$ , definida como o inverso da energia de ativação aparente, pode ser interpretada como a medida da tendência de uma reação ocorrer. Esta função fornece uma nova perspectiva sobre o gráfico de *Arrhenius*, pois permite linearizar curvaturas de processos não-*Arrhenius* e auxiliar na percepção de mudanças de regimes durante os processos que não são percebidas no gráfico usual. O relaxamento do carbonato de propileno é um processo não *Arrhenius* que apresenta uma curvatura acentuada no gráfico de *Arrhenius*, mas ao ser visualizado no gráfico da transitividade, apresentou um comportamento linear com mudança de inclinação à temperatura de 198K: uma quebra de regime que provoca uma mudança na dinâmica do processo. A função transitividade, ajustada para cada regime, mostrou que, para temperaturas abaixo de 198K, o processo apresenta maior altura de barreira e menor valor do parâmetro de deformação. Acompanhando o modelo matemático proposto, foi desenvolvido o código *Transitivity*, que sistematiza ferramentas desenvolvidas ao longo dos últimos 10 anos para lidar com a cinética de processos não *Arrhenius*, realizando a estimativa da constante cinética teórica e ajustes de modelos cinéticos a partir de dados experimentais para obtenção de parâmetros fenomenológicos. Além disso, o código *Transitivity* foi utilizado para realizar uma análise completa do diagrama do perfil de energia livre de *Gibbs* na síntese de chalconas elucidando detalhes acerca deste mecanismo.

A compreensão acerca da influência da temperatura na velocidade de processos é de extrema importância também em processos biológicos, principalmente para prever o impacto de mudanças climáticas e compreender o funcionamento do metabolismo de seres vivos. O coeficiente da temperatura  $Q_{10}$  é a ferramenta frequentemente utilizada para observar estes fenômenos, entretanto ela falha quando submetida a um largo intervalo de temperatura. No intuito de fornecer

uma melhor abordagem para esta análise, foi aqui proposto o modelo para o coeficiente generalizado da temperatura  $Q_{d10}$ . O modelo proposto foi aplicado a diversos casos da literatura recente e a qualidade dos ajustes reforçou a hipótese inicial de que as constantes cinéticas utilizadas devem ser descritas por exponenciais deformadas ao invés das exponenciais convencionais, como estabelecido pelo modelo padrão de *Arrhenius*. Além disso, o modelo padrão  $Q_{10}$  é um caso particular do modelo proposto e valida sua justificativa geral: em casos extremos onde  $d = 0$  a constante cinética do processo varia de forma constante quando a temperatura é aumentada em 10 unidades, como descrito pelo modelo padrão. No entanto, em casos gerais ( $d \neq 0$  e não *Arrhenius*) a razão entre as taxas pode obter qualquer valor positivo. Deste modo, o modelo generalizado  $Q_{d10}$  proposto foi validado com sucesso utilizando dados experimentais, sugerindo a existência de uma larga variedade de aplicações para o  $Q_{d10}$  na descrição de constantes cinéticas dependentes da temperatura, principalmente em processos biológicos. Além disso, o código  $Q_{d10}$  – GSA foi desenvolvido e estima parâmetros fenomenológicos relacionados as formulações de *Arrhenius* e Aquilanti-Mundim a partir de dados experimentais, traçando assim um paralelo entre o modelo usual  $Q_{10}$  e o modelo aqui proposto  $Q_{d10}$ .

É importante notar que os modelos matemáticos aqui propostos tem duas bases principais: o uso consistente de exponenciais deformadas ( $d$ ) em suas formulações e o uso de algoritmos de *machine learning* na compreensão e desenvolvimento destes modelos. O *machine learning* é a principal área da inteligência artificial e seus desafios recentes envolvem a utilização de algoritmos *deep learning* na resolução de problemas com soluções intuitivas, como é o caso da análise de técnicas espectroscópicas na química. Neste sentido, foram aqui propostos modelos de *deep learning* que utilizam e aprendem com dados espectrais de infravermelho e espectrometria de massa, juntamente ao desenvolvimento da plataforma *SpectralAI* para utilização destes modelos. A CNN ResNet34 e o modelo tabular proposto – que utilizam respectivamente imagens e dados espectrais brutos – apresentaram ambos um bom desempenho na classificação de características estruturais moleculares utilizando dados de infravermelho (IR), com valores de *F1-Score* entre 0,83 e 0,88 e acurácias de 96%. A pequena melhora nas métricas pela utilização de espectros concatenados de infravermelho com espectros de massa (IR

+ MS) não foi suficiente para definir esta metodologia como a mais apropriada. Já os modelos desenvolvidos para reconstrução de espectros teóricos, apesar de demonstrarem um certo aprendizado reproduzindo a tendência de alguns picos do espectro, não apresentaram resultados satisfatórios, com valores de erro quadrático médio 0,005 e 0,007 para os espectros de MS e IR, respectivamente. Dados espectrais possuem padrões altamente complexos. Por isso, acredita-se que o tamanho do conjunto de dados foi um fator limitante no aprendizado destes modelos, uma vez que o espectro de IR, por exemplo, apresenta 4000 valores de intensidades, enquanto nosso conjunto de dados contempla pouco mais de 1600 compostos. Acredita-se que com um banco de dados suficientemente grande os modelos poderão desempenhar de modo satisfatório, tanto na tarefa de análise do espectro experimental quanto na geração de espectros teóricos, além de aumentarem seu domínio de aplicabilidade. Com isso, conclui-se que o desenvolvimento destes modelos sugere que algoritmos de *deep learning* baseados em redes neurais são adequados para a busca de novas técnicas de análises espectrais e representam um passo no sentido da criação de técnicas de análises espectrais autônomas.

Por fim, os modelos aqui propostos foram utilizados em uma série de aplicações bem-sucedidas, desde descrições fenomenológicas até descrições de primeiros princípios de cinética química pura e aplicada a ciência dos materiais, sempre com o auxílio de algoritmos de inteligência artificial. As plataformas *Transitivity*, *Q<sub>d10</sub> – GSA* e *SpectralAI* aqui desenvolvidas, apresentam interfaces bastante intuitivas e amigáveis ao usuário, contribuindo com a disseminação dos conceitos e modelos propostos. Além da importância dos modelos na caracterização de fenômenos químicos e biológicos, o autor espera que as plataformas sejam utilizadas como poderosas ferramentas de pesquisa e ferramentas didáticas para o ensino de termodinâmica, cinética de processos, espectroscopia química, inteligência artificial e assuntos relacionados.

Como perspectiva deste trabalho, pretende-se desenvolver aplicações *web* para o *Transitivity* e *Q<sub>d10</sub> – GSA* que substituirão os *softwares* atuais a fim de melhorar ainda mais a experiência do usuário e contribuir para investigações futuras utilizando os modelos aqui propostos, além de consolidar o desenvolvimento de *software* como uma forte característica de nossos grupos de pesquisa. Quanto aos

modelos de *deep learning* aqui desenvolvidos para análise espectroscópica, sabe-se que o banco de dados é um fator determinante para o desempenho desses modelos, e, portanto, o autor pretende seguir no sentido de incrementar o banco de dados atual, além de buscar novos conjuntos de dados de técnicas espectroscópicas como: ressonância magnética nuclear; difração de raios-x, espectroscopia eletrônica; dentre outros. Com isso, espera-se que a utilização de algoritmos de inteligência artificial se torne uma característica dos nossos grupos de pesquisa, servindo como ferramentas em pesquisas e investigações futuras.

## REFERÊNCIAS BIBLIOGRÁFICAS

1. Turing AM. Computing Machinery and Intelligence. *Mind*. 1950;(49):433-460.
2. Dundas J, Chik D. Implementing Human-like Intuition Mechanism in Artificial Intelligence. Published online 2011. <http://arxiv.org/abs/1106.5917>
3. Aires-de-Sousa J, Hemmer MC, Gasteiger J. Prediction of 1H NMR chemical shifts using neural networks. *Anal Chem*. 2002;74(1):80-90. doi:10.1021/ac010737m
4. Zupan J, Gasteiger J. Neural Networks in Chemistry and Drug Design. Published online 1999:400. doi:3-527-29779-0
5. Kwon S, Bae H, Jo J, Yoon S. Comprehensive ensemble in QSAR prediction for drug discovery. *BMC Bioinformatics*. 2019;20(1):1-12. doi:10.1186/s12859-019-3135-4
6. Balaban AT. Neural Networks in QSAR and Drug Design. *J Chem Inf Comput Sci*. 1997;37(3).
7. Mundim KC, Tsallis C. Geometry optimization and conformational analysis through generalized simulated annealing. *Int J Quantum Chem*. 1996;58(4):373-381. doi:10.1002/(sici)1097-461x(1996)58:4<373::aid-qua6>3.0.co;2-v
8. Behler J. Neural network potential-energy surfaces in chemistry: A tool for large-scale simulations. *Physical Chemistry Chemical Physics*. 2011;13(40):17930-17955. doi:10.1039/c1cp21668f
9. Ramsundar B, Riley P, Webster D, Konerding D, Edu KS, Edu PS. Massively Multitask Networks for Drug Discovery. 2015;(Icml).
10. Ma J, Sheridan RP, Liaw A, Dahl GE, Svetnik V. Deep neural nets as a method for quantitative structure-activity relationships. *J Chem Inf Model*. 2015;55(2):263-274. doi:10.1021/ci500747n
11. Hamza H, Salim N, Nasser M, Saeed F. AtomNet: A Deep Learning Neural Network for Bioactivity Prediction in Structure-based Drug Discovery. Published online 2020:21-37. doi:10.5121/csit.2020.100203
12. Lusci A, Pollastri G, Baldi P. Deep architectures and deep learning in chemoinformatics: The prediction of aqueous solubility for drug-like molecules. *J Chem Inf Model*. 2013;53(7):1563-1575. doi:10.1021/ci400187y
13. Mobley DL, Wymer KL, Lim NM, Guthrie JP. Blind prediction of solvation free energies from the SAMPL4 challenge. *J Comput Aided Mol Des*. 2014;28(3):135-150. doi:10.1007/s10822-014-9718-2
14. Delaney JS. ESOL: Estimating aqueous solubility directly from molecular structure. *J Chem Inf Comput Sci*. 2004;44(3):1000-1005. doi:10.1021/ci034243x
15. Mobley DL, Guthrie JP. FreeSolv: A database of experimental and calculated hydration free energies, with input files. *J Comput Aided Mol Des*. 2014;28(7):711-720. doi:10.1007/s10822-014-9747-x

16. Rupp M, Tkatchenko A, Müller KR, Von Lilienfeld OA. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys Rev Lett*. 2012;108(5):1-5. doi:10.1103/PhysRevLett.108.058301
17. Montavon G, Rupp M, Gobre V, et al. Machine learning of molecular electronic properties in chemical compound space. *New J Phys*. 2013;15:0-16. doi:10.1088/1367-2630/15/9/095003
18. McGibbon RT, Taube AG, Donchev AG, et al. Improving the accuracy of Møller-Plesset perturbation theory with neural networks. *Journal of Chemical Physics*. 2017;147(16). doi:10.1063/1.4986081
19. Schütt KT, Arbabzadah F, Chmiela S, Müller KR, Tkatchenko A. Quantum-chemical insights from deep tensor neural networks. *Nat Commun*. 2017;8(0):1-21. doi:10.1038/ncomms13890
20. Aquilanti V, Coutinho ND, Carvalho-Silva VH. Kinetics of low-temperature transitions and a reaction rate theory from non-equilibrium distributions. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 2017;375(2092). doi:10.1098/rsta.2016.0201
21. De Oliveira GR, De Oliveira HCB, Silva WA, Da Silva VHC, Sabino JR, Martins FT. Structure and theoretical approaches to a chalcone derivative. *Struct Chem*. 2012;23(6):1667-1676. doi:10.1007/s11224-012-9972-7
22. Rezende MVCS, Coutinho ND, Palazzetti F, Lombardi A, Carvalho-Silva VH. Nucleophilic substitution vs elimination reaction of bisulfide ions with substituted methanes: exploration of chiral selectivity by stereodirectional first-principles dynamics and transition state theory. *J Mol Model*. 2019;25(8). doi:10.1007/s00894-019-4126-0
23. Sanches-Neto FO, Coutinho ND, Palazzetti F, Carvalho-Silva VH. Temperature dependence of rate constants for the H(D) + CH<sub>4</sub> reaction in gas and aqueous phase: deformed Transition-State Theory study including quantum tunneling and diffusion effects. *Struct Chem*. 2020;31(2):609-617. doi:10.1007/s11224-019-01437-3
24. Sallum LDO, Napolitano HB, Carvalho PDS, et al. Effect of the Methanol Molecule on the Stabilization of C<sub>18</sub>H<sub>18</sub>O<sub>4</sub> Crystal: Combined Theoretical and Structural Investigation. Published online 2014.
25. Carvalho-Silva V.H., Vaz E.C., Coutinho N.D., Kobayashi H., Kobayashi Y., Kasai T., Palazzetti F., Lombardi A. A v. The Increase of the Reactivity of Molecular Hydrogen with Hydroxyl Radical from the Gas Phase versus an Aqueous Environment: Quantum Chemistry and Transition State-Theory Calculations. *Lecture Notes in Computer Science*. 2019.
26. Aquilanti V, Borges EP, Coutinho ND, Mundim KC, Carvalho-Silva VH. From statistical thermodynamics to molecular kinetics: the change, the chance and the choice. *Rendiconti Lincei*. 2018;29(4):787-802. doi:10.1007/s12210-018-0749-9
27. COUTINHO, N. D.; SILVA, Y. S.; FAZIO, D.; CAVALLI, S.; CARVALHO-SILVA, VALTER H.; AQUILANTI V. Chemical Kinetics under Extreme

- Conditions: Exact, Phenomenological and First-Principles Computational Approaches. Chemical Kinetics under Extreme Conditions: Exact, Phenomenological and First-Principles Computational Approaches. No Title. *ACCADEMIA NAZIONALE DELLE SCIENZE DETTA DEI XL*. Published online 2019.
28. Carvalho-Silva VH, Coutinho ND, Aquilanti V. From the kinetic theory of gases to the kinetics of rate processes: On the verge of the thermodynamic and kinetic limits. *Molecules*. 2020;25(9). doi:10.3390/molecules25092098
  29. Santin LG, Toledo EM, Carvalho-Silva VH, Camargo AJ, Gargano R, Oliveira SS. Methanol Solvation Effect on the Proton Rearrangement of Curcumin's Enol Forms: An Ab Initio Molecular Dynamics and Electronic Structure Viewpoint. *The Journal of Physical Chemistry C*. Published online 2016:acs.jpcc.6b02393. doi:10.1021/acs.jpcc.6b02393
  30. Aquilanti V, Mundim KC, Elango M, Kleijn S, Kasai T. Temperature dependence of chemical and biophysical rate processes: Phenomenological approach to deviations from Arrhenius law. *Chem Phys Lett*. 2010;498(1-3):209-213. doi:10.1016/j.cplett.2010.08.035
  31. Machado HG, Sanches-Neto FO, Coutinho ND, Mundim KC, Palazzetti F, Carvalho-Silva VH. "Transitivity": A code for computing kinetic and related parameters in chemical transformations and transport phenomena. *Molecules*. 2019;24(19). doi:10.3390/molecules24193478
  32. Mundim KC, Baraldi S, Machado HG, Vieira FMC. Temperature coefficient (Q<sub>10</sub>) and its applications in biological systems: Beyond the Arrhenius theory. *Ecol Modell*. 2020;431(December 2019). doi:10.1016/j.ecolmodel.2020.109127
  33. Lopes TO, Machado DFS, Oliveira HCB. UV-vis Spectrum Patronus. Published online 2017.
  34. Atkinson R, Carter WPL. Kinetics and Mechanisms of the Gas-Phase Reactions of the hydroxyl radical with Organic Compounds under Atmospheric Conditions. *Chem Rev*. 1986;84(5):437-470. doi:10.1021/cr00063a002
  35. Limbach HH, Lopez JM, Kohen A. Arrhenius curves of hydrogen transfers: Tunnel effects, isotope effects and effects of pre-equilibria. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2006;361(1472):1399-1415. doi:10.1098/rstb.2006.1872
  36. Smith IWM. The temperature-dependence of elementary reaction rates: Beyond Arrhenius. *Chem Soc Rev*. 2008;37(4):812-826. doi:10.1039/b704257b
  37. Sims IR. Low-Temperature Reactions: Tunnelling in Space. *Nat Chem*. 2013;5(9):734-736. doi:10.1038/nchem.1736
  38. Peleg M, Normand MD, Corradini MG. The Arrhenius equation revisited. *Crit Rev Food Sci Nutr*. 2012;52(9):830-851. doi:10.1080/10408398.2012.667460
  39. Darrington RT, Jiao J. Rapid and Accurate Prediction of Degradant Formation Rates in Pharmaceutical Formulations Using High-Performance Liquid Chromatography-Mass Spectrometry. *J Pharm Sci*. 2004;93(4):838-846. doi:10.1002/jps.20006

40. Giordano D, Russell JK. Towards a structural model for the viscosity of geological melts. *Earth Planet Sci Lett.* 2018;501:202-212. doi:10.1016/j.epsl.2018.08.031
41. Klinman JP, Kohen A. Hydrogen tunneling links protein dynamics to enzyme catalysis. *Annu Rev Biochem.* 2013;82:471-496. doi:10.1146/annurev-biochem-051710-133623
42. Warshel A, Bora RP. Perspective: Defining and quantifying the role of dynamics in enzyme catalysis. *Journal of Chemical Physics.* 2016;144(18). doi:10.1063/1.4947037
43. Kooij DM. Über die Zersetzung des gasförmigen Phosphorwasserstoffs. *Zeitschrift für Phys Chemie.* 1893;12:155-161. doi:https://doi.org/10.1515/zpch-1893-1214
44. Vogel H. Das temperature-abhängigkeitsgesetz der viskosität von flüssigkeiten. *Phys Z.* 1921;22:645-646.
45. Fulcher GS. Analysis of Recent Measurements of the Viscosity of Glasses. *Journal of the American Ceramic Society.* 1925;8(6):339-355. doi:10.1111/j.1151-2916.1925.tb16731.x
46. Tammann G, Hesse W. Die Abhängigkeit der Viskosität von der Temperatur bei unterkühlten Flüssigkeiten. *Z Anorg Allg Chem.* 1926;156(1):245-257. doi:10.1002/zaac.19261560121
47. Nakamura K, Takayanagi T, Sato S. A modified arrhenius equation. *Chem Phys Lett.* 1989;160(3):295-298. doi:10.1016/0009-2614(89)87599-2
48. Isaacson AD, Truhlar DG, Rai SN, et al. POLYRATE: A general computer program for variational transition state theory and semiclassical tunneling calculations of chemical reaction rates. *Comput Phys Commun.* 1987;47(1):91-102. doi:10.1016/0010-4655(87)90069-5
49. Duncan WT, Bell RL, Truong TN. TheRate: Program for ab initio direct dynamics calculations of thermal and vibrational-state-selected rate constants. *J Comput Chem.* 1998;19(9):1039-1052. doi:10.1002/(SICI)1096-987X(19980715)19:9<1039::AID-JCC5>3.0.CO;2-R
50. Barker JR. Multiple-well, multiple-path unimolecular reaction systems. I. MultiWell computer program suite. *Int J Chem Kinet.* 2001;33(4):232-245. doi:10.1002/kin.1017
51. Ghysels A, Verstraelen T, Hemelsoet K, Waroquier M, Van Speybroeck V. TAMkin: A versatile package for vibrational analysis and chemical kinetics. *J Chem Inf Model.* 2010;50(9):1736-1750. doi:10.1021/ci100099g
52. Glowacki DR, Liang CH, Morley C, Pilling MJ, Robertson SH. MESMER: An open-source master equation solver for Multi-Energy well reactions. *Journal of Physical Chemistry A.* 2012;116(38):9545-9560. doi:10.1021/jp3051033
53. Gao CW, Allen JW, Green WH, West RH. Reaction Mechanism Generator: Automatic construction of chemical kinetic mechanisms. *Comput Phys Commun.* 2016;203:212-225. doi:10.1016/j.cpc.2016.02.013



54. Euclides HO, Patricia PR. APUAMA: a software tool for reaction rate calculations. *J Mol Model*. 2017;23(6). doi:10.1007/s00894-017-3337-5
55. Canneaux S, Bohr F, Henon E. KiSThelP: A program to predict thermodynamic properties and rate constants from quantum chemistry results. *J Comput Chem*. 2014;35(1):82-93. doi:10.1002/jcc.23470
56. Coppola CM, Kazandjian M V. Matrix formulation of the energy exchange problem of multi-level systems and the code FRIGUS. *Rendiconti Lincei*. 2019;30(4):707-714. doi:10.1007/s12210-019-00849-x
57. Dzib E, Cabellos JL, Ortíz-Chi F, Pan S, Galano A, Merino G. Eyringpy: A program for computing rate constants in the gas phase and in solution. *Int J Quantum Chem*. 2019;119(2):11-13. doi:10.1002/qua.25686
58. Svante Arrhenius. Zeitschrift für Physikalische Chemie. *Phys Chem*. 1889;4(226).
59. Trautz M. Das Gesetz der Reaktionsgeschwindigkeit und der Gleichgewichte in Gasen. *Z Anorg Allg Chem*. 1916;96(1):1-28. doi:10.1002/zaac.19160960102
60. Eyring H. The Activated Complex in Chemical Reactions. *J Chem Phys*. 1935;3(2):107-115. doi:10.1063/1.1749604
61. Carvalho-Silva VH, Aquilanti V, de Oliveira HCB, Mundim KC. Deformed transition-state theory: Deviation from Arrhenius behavior and application to bimolecular hydrogen transfer reaction rates in the tunneling regime. *J Comput Chem*. 2017;38(3):178-188. doi:10.1002/jcc.24529
62. Claudino D, Gargano R, Carvalho-Silva VH, E Silva GM, da Cunha WF. Investigation of the Abstraction and Dissociation Mechanism in the Nitrogen Trifluoride Channels: Combined Post-Hartree-Fock and Transition State Theory Approaches. *Journal of Physical Chemistry A*. 2016;120(28):5464-5473. doi:10.1021/acs.jpca.6b04947
63. Leonhard Euler. *Introductio In Analysin Infinitorum*. Vol 1. Bousquet; 1748.
64. Ratti C. *Hot Air and Freeze-Drying of High-Value Foods: A Review.*; 2001. doi:https://doi.org/10.1016/S0260-8774(00)00228-4
65. Dalziel K. *Kinetic Studies of Liver Alcohol Dehydrogenase*. Vol 80. Theorell, H. & Nygaard, A. P; 1962.
66. Smith RS, Kay BD. Breaking through the glass ceiling: Recent experimental approaches to probe the properties of supercooled liquids near the glass transition. *Journal of Physical Chemistry Letters*. 2012;3(6):725-730. doi:10.1021/jz201710z
67. Debenedetti PG, Stillinger FH. *Supercooled Liquids and the Glass Transition.*; 2001. doi:https://doi.org/10.1038/35065704
68. Silva VHC, Aquilanti V, Oliveira HCB de, Mundim KC. Uniform description of non-Arrhenius temperature dependence of reaction rates , and a heuristic criterion for quantum tunneling vs classical non-extensive distribution. *Chem Phys Lett*. 2013;590:201-207. doi:10.1016/j.cplett.2013.10.051
69. Tsallis C. Possible Generalization of Boltzmann-Gibbs Statistics. *J Stat Phys*. 1988;52(1):479-487. doi:10.1007/BF01016429

70. Bell RP, A PRSL. Quantum mechanical effects in reactions involving hydrogen. *Proc R Soc Lond A Math Phys Sci.* 1935;148(864):241-250. doi:10.1098/rspa.1935.0016
71. Cavalli S, Aquilanti V, Mundim KC, de Fazio D. Theoretical reaction kinetics astride the transition between moderate and deep tunneling regimes: The F + HD case. *Journal of Physical Chemistry A.* 2014;118(33):6632-6641. doi:10.1021/jp503463w
72. Bell RP. The Tunnel Effect Correction for Parabolic Potential Barriers. *Transactions of the Faraday Society.* Published online 1959:1-4.
73. Bell RP. *The Tunnel Effect in Chemistry.* Chapman and Hall; 1980.
74. Christov SG. The Characteristic (Crossover) Temperature in the Theory of Thermally Activated Tunneling Processes. Published online 1997:109-147. doi:10.1007/978-94-011-4894-8\_7
75. Skodje RT, Truhlar DG. Parabolic tunneling calculations. *Journal of Physical Chemistry.* 1981;85(6):624-628. doi:10.1021/j150606a003
76. Collins FC, Kimball GE. Diffusion-controlled reaction rates. *J Colloid Sci.* 1949;4(4):425-437. doi:10.1016/0095-8522(49)90023-9
77. Collins FC, Kimball GE. Diffusion-Controlled Reactions in Liquid Solutions. *Ind Eng Chem.* 1949;41(11):2551-2553. doi:10.1021/ie50479a040
78. Kramers HA. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica.* 1940;7(4):284-304. doi:10.1016/S0031-8914(40)90098-2
79. Onsager L. Electric Moments of Molecules in Liquids. *J Am Chem Soc.* 1936;58(8):1486-1493. doi:10.1021/ja01299a050
80. Wah MW, Wiberg KB, Frisch MJ. Solvent Effects. 3. Tautomeric Equilibria of Formamide and 2-Pyridone in the Gas Phase and Solution. An ab Initio SCRF Study. *J Am Chem Soc.* 1992;114(5):1645-1652. doi:10.1021/ja00031a017
81. Henriksen NE, Hansen FY. *Theories of Molecular Reaction Dynamics: The Microscopic Foundation of Chemical Kinetics.* Oxford University Press; 2008. doi:10.1093/acprof:oso/9780199203864.001.0001
82. Von Smoluchowski M. Drei Vortrage uber Diffusion. Brownsche Bewegung und Koagulation von Kolloidteilchen. *Z Phys.* 1916;17:557-585.
83. Eigen M. Proton Transfer, Acid-Base Catalysis, and Enzymatic Hydrolysis. Part I: ELEMENTARY PROCESSES. *Angewandte Chemie International Edition in English.* 1964;3(1):1-19. doi:10.1002/anie.196400011
84. Carvalho-Silva VH, Coutinho ND, Aquilanti V. Temperature dependence of rate processes beyond Arrhenius and Eyring: Activation and transitivity. *Front Chem.* 2019;7(MAY). doi:10.3389/fchem.2019.00380
85. Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by simulated annealing. *Science (1979).* 1983;220(4598):671-680. doi:10.1126/science.220.4598.671
86. Ceperley D, Alder B. Quantum Monte Carlo. *Science (1979).* Published online 1986.

87. Szu H. Fast Simulated Annealing\*. *American Institute of Physics*. 1986;(2):420-425.
88. Andrade MD de, Mundim KC, Malbouisson LAC. Convergence of the Generalized Simulated Annealing Method With Independent Parameters for the Acceptance Probability, Visitation Distribution, and Temperature Functions. *Int J Quantum Chem*. 2008;108:2392-2397. doi:10.1002/qua
89. Laidler KJ. A glossary of terms used in chemical kinetics, including reaction dynamics (IUPAC recommendations 1996). *Pure and Applied Chemistry*. 1996;68(1):149-192. doi:10.1351/pac199668010149
90. A. R. Ravishankara, Nicovich JM, Thompson RL, Tully FP. Kinetic Study of the Reaction of OH with H<sub>2</sub> and D<sub>2</sub> from 230 to 1050K. *J Phys Chem*. 1981;85(17):2498-2503. doi:https://doi.org/10.1021/j150617a018
91. Kohen A, Cannio R, Bartolucci S, Klinman JP. Enzyme dynamics and hydrogen tunnelling in a thermophilic alcohol dehydrogenase. *Nature*. 1999;399(6735):496-499. doi:10.1038/20981
92. Liang ZX, Tsigos I, Bouriotis V, Klinman JP. Impact of protein flexibility on hydride-transfer parameters in thermophilic and psychrophilic alcohol dehydrogenases. *J Am Chem Soc*. 2004;126(31):9500-9501. doi:10.1021/ja047087z
93. Truhlar DG, Kohen A. Convex Arrhenius plots and their interpretation. *Proc Natl Acad Sci U S A*. 2001;98(3):848-851. doi:10.1073/pnas.98.3.848
94. Coutinho ND, Aquilanti V, Silva VHC, Camargo AJ, Mundim KC, De Oliveira HCB. Stereodirectional Origin of anti-Arrhenius Kinetics for a Tetraatomic Hydrogen Exchange Reaction: Born-Oppenheimer Molecular Dynamics for OH + HBr. *Journal of Physical Chemistry A*. 2016;120(27):5408-5417. doi:10.1021/acs.jpca.6b03958
95. Oliveira-filho AGS De, Ornellas FR, Bowman JM. Quasiclassical Trajectory Calculations of the Rate Constant of the OH. Published online 2014.
96. Coutinho ND, Silva VHC, De Oliveira HCB, Camargo AJ, Mundim KC, Aquilanti V. Stereodynamical origin of anti-arrhenius kinetics: Negative activation energy and roaming for a four-atom reaction. *Journal of Physical Chemistry Letters*. 2015;6(9):1553-1558. doi:10.1021/acs.jpcllett.5b00384
97. Zuniga-Hansen N, Silbert LE, Calbi MM. Breakdown of kinetic compensation effect in physical desorption. *Phys Rev E*. 2018;98(3):1-10. doi:10.1103/PhysRevE.98.032128
98. Sims IR, Smith IWM, Clary DC, Bocherel P, Rowe BR. Ultra-low temperature kinetics of neutral-neutral reactions: New experimental and theoretical results for OH+HBr between 295 and 23 K. *J Chem Phys*. 1994;101(2):1748-1751. doi:10.1063/1.467733
99. Souletie J, Tholence JL. Critical slowing down in spin glasses and other glasses: Fulcher versus power law. *Phys Rev B*. 1985;32(1):516-519. doi:10.1103/PhysRevB.32.516

100. Stickel F, Fischer EW, Richert R. Dynamics of glass-forming liquids. II. Detailed comparison of dielectric relaxation, de-conductivity, and viscosity data. *Journal of Chemical Physics*. 1996;104(5):2043-2055. doi:10.1063/1.470961
101. Drozd-Rzoska A. Universal behavior of the apparent fragility in ultraslow glass forming systems. *Sci Rep*. 2019;9(1):1-9. doi:10.1038/s41598-019-42927-y
102. Frisch MJ, Trucks GW, Schlegel HB, et al. *Gaussian 09, Revision A.02*. Wallingford CT; 2016.
103. Coutinho ND, Sanches-Neto FO, Carvalho-Silva VH, de Oliveira HCB, Ribeiro LA, Aquilanti V. Kinetics of the OH+HCl→H<sub>2</sub>O+Cl reaction: Rate determining roles of stereodynamics and roaming and of quantum tunneling. *J Comput Chem*. 2018;39(30):2508-2516. doi:10.1002/jcc.25597
104. Ravishankara AR, Wine PH, Wells JR, Thompson RL. Kinetic study of the reaction of OH with HCl from 240–1055 K. *Int J Chem Kinet*. 1985;17(12):1281-1297. doi:10.1002/kin.550171206
105. Hickel B, Sehested K. Reaction of hydroxyl radicals with ammonia in liquid water at elevated temperatures. *International Journal of Radiation Applications and Instrumentation Part*. 1992;39(4):355-357. doi:10.1016/1359-0197(92)90244-A
106. Men'kin VB, Makarov IE, Pikaev AK. Pulse radiolysis study of reaction rates of OH and O- radicals with ammonia in aqueous solutions. *High Energy Chemistry (English Translation)*. 1989;22(5):333-336.
107. Neta P, Maruthamuthu P, Carton PM, Fessenden RW. Formation and reactivity of the amino radical. *Journal of Physical Chemistry*. 1978;82(17):1875-1878. doi:10.1021/j100506a004
108. Coutinho ND, Machado HG, Carvalho-Silva VH, da Silva WA. Topography of the free energy landscape of Claisen–Schmidt condensation: solvent and temperature effects on the rate-controlling step. *Phys Chem Chem Phys*. 2021;23:6738-6745. doi:10.1039/D0CP05659F
109. Claisen L, Claparède A. Condensationen von Ketonen mit Aldehyden. *Ber Dtsch Chem Ges*. 1881;14(1):2460-2468.
110. Schmidt JG. Ueber die Einwirkung von Aceton auf Furfurol und auf Bittermandelöl bei Gegenwart von Alkalilauge. *Berichte der deutschen chemischen Gesellschaft*. 1881;14(1):1459-1461. doi:10.1002/cber.188101401306
111. Nielsen AT, Houlihan WJ. The Aldol Condensation. *Organic Reactions*. 2011;16:1-438. doi:10.1002/0471264180.or016.01
112. Mukaiyama T. The Directed Aldol Reaction. *Organic Reactions*. 1982;28:203-331. doi:10.1002/0471264180.or028.03
113. Guthrie JP. Rate Equilibrium COrelations for the Aldol Condensation: An Analysis in Terms of Marcus Theory. *American Chemical Society*. 1991;847(4):7249-7255.
114. Dalessandro EV, Collin HP, Guimarães LGL, Valle MS, Pliego JR. Mechanism of the Piperidine-Catalyzed Knoevenagel Condensation Reaction in Methanol:

- The Role of Iminium and Enolate Ions. *Journal of Physical Chemistry B*. 2017;121(20):5300-5307. doi:10.1021/acs.jpcc.7b03191
115. Dalessandro E V., Collin HP, Valle MS, Pliego JR. Mechanism and free energy profile of base-catalyzed Knoevenagel condensation reaction. *RSC Adv*. 2016;6(63):57803-57810. doi:10.1039/c6ra10393f
  116. Evans P, Coombs E. Condensations of Carbonyl Compounds. A Kinetic Study of the Reaction of Acetophenone with benzaldehyde. Published online 1940:4-6. doi:10.1039/JR9400001295
  117. Perrin CL, Chang KL. The Complete Mechanism of an Aldol Condensation. *Journal of Organic Chemistry*. 2016;81(13):5631-5635. doi:10.1021/acs.joc.6b00959
  118. Mak KKW, Chan W fat, Lung K ying, Lam W yee, Ng W cheong, Lee S fung. Probing the Rate-Determining Step of the Claisen – Schmidt Condensation by Competition Reactions. *J Chem Educ*. 2007;84(11):1819-1821.
  119. Gasull EI, Silber JJ, Blanco SE, Tomas F, Ferretti FH. A theoretical and experimental study of the formation mechanism of 4-X-chalcones by the Claisen–Schmidt reaction. *Journal of Molecular Structure: THEOCHEM*. 2000;503(3):131-144. doi:10.1016/S0166-1280(99)00256-0
  120. Guida A, Lhouty MH, Tichit D, Figueras F, Geneste P. Hydrotalcites as base catalysts. Kinetics of Claisen-Schmidt condensation, intramolecular condensation of acetonylacetone and synthesis of chalcone. *Appl Catal A Gen*. 1997;164(1-2):251-264. doi:10.1016/S0926-860X(97)00175-0
  121. Richard JP, Nagorski RW. Mechanistic imperatives for catalysis of aldol addition reactions: Partitioning of the enolate intermediate between reaction with bronsted acids and the carbonyl group. *J Am Chem Soc*. 1999;121(20):4763-4770. doi:10.1021/ja9900297
  122. Zhao GM, Sun WJ, Zhang HL, Li DL, Yang X. DFT study on selective autocatalyzed  $\alpha$ -alkylation of ketones with alcohols. *J Catal*. 2019;373:126-138. doi:10.1016/j.jcat.2019.03.033
  123. Zhao Y, Zhu X, Wang H, Han J, Mei D, Ge Q. Aqueous Phase Aldol Condensation of Formaldehyde and Acetone on Anatase TiO<sub>2</sub>(101) Surface: A Theoretical Investigation. *ChemCatChem*. 2020;12(4):1220-1229. doi:10.1002/cctc.201901736
  124. Ley D, Gerbig D, Schreiner PR. Tunnelling control of chemical reactions - The organic chemist's perspective. *Org Biomol Chem*. 2012;10(19):3781-3790. doi:10.1039/c2ob07170c
  125. Castro C, Karney WL. Heavy-Atom Tunneling in Organic Reactions. *Angewandte Chemie - International Edition*. 2020;59(22):8355-8366. doi:10.1002/anie.201914943
  126. Schreiner PR. Tunneling Control of Chemical Reactions: The Third Reactivity Paradigm. *J Am Chem Soc*. 2017;139(43):15276-15283. doi:10.1021/jacs.7b06035

127. Grimme S. Semiempirical hybrid density functional with perturbative second-order correlation Semiempirical hybrid density functional with perturbative second-order correlation. *J Chem Phys.* 2006;124(034108). doi:10.1063/1.2148954
128. Krishnan R, Binkley JS, Seeger R, Pople JA. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *J Chem Phys.* 1980;72(1):650-654. doi:10.1063/1.438955
129. Marenich A V., Cramer CJ, Truhlar DG. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *Journal of Physical Chemistry B.* 2009;113(18):6378-6396. doi:10.1021/jp810292n
130. Grimme S. Density functional theory with London dispersion corrections. *Computational Molecular Science.* 2011;1(April):211-228. doi:10.1002/wcms.30
131. Noyce DS, Pryor W a. Carbonyl reactions. I. Kinetics and mechanism of the acid-catalyzed aldol condensation of benzaldehyde and acetophenone. *Journal of american chemical society.* 1955;77(9):1397-1401. doi:10.1021/ja01611a001
132. Silva WA, Andrade CKZ, Napolitano HB, et al. Biological and structure-activity evaluation of chalcone derivatives against bacteria and fungi. *J Braz Chem Soc.* 2013;24(1):133-144. doi:10.1590/S0103-50532013000100018
133. Gomes KS, da Costa-Silva TA, Oliveira IH, et al. Structure-activity relationship study of antitrypanosomal chalcone derivatives using multivariate analysis. *Bioorg Med Chem Lett.* 2019;29(12):1459-1462. doi:10.1016/j.bmcl.2019.04.020
134. Guthrie JP, Cooper KJ, Cossar J. The retroaldol reaction of cinnamaldehyde. *Can J Chem.* 1984;62(8):1441-1451. doi:10.1139/v84-243
135. Iglesias E. Solvent effects versus concentration effects in determining rates of base-catalyzed keto-enol tautomerization. *New Journal of Chemistry.* 2005;29(4):625-632. doi:10.1039/b415220d
136. Perrin CL. Are short, low-barrier hydrogen bonds unusually strong? *Acc Chem Res.* 2010;43(12):1550-1557. doi:10.1021/ar100097j
137. Alagona G, Ghio C, Nagy PI. The catalytic effect of water on the keto-enol tautomerism. Pyruvate and acetylacetone: A computational challenge. *Physical Chemistry Chemical Physics.* 2010;12(35):10173-10188. doi:10.1039/c003999c
138. Bunnett JF. The Mechanism of Bimolecular  $\beta$ -Elimination Reactions. *Angewandte Chemie International Edition in English.* 1962;1(5):225-235. doi:10.1002/anie.196202251
139. Perrin CL. Symmetry of hydrogen bonds in solution. *Pure and Applied Chemistry.* 2009;81(4):571-583. doi:10.1351/PAC-CON-08-08-14
140. Santin LG, Toledo EM, Carvalho-Silva VH, Camargo AJ, Gargano R, Oliveira SS. Methanol Solvation Effect on the Proton Rearrangement of Curcumin's

- Enol Forms: An Ab Initio Molecular Dynamics and Electronic Structure Viewpoint. *Journal of Physical Chemistry C*. 2016;120(36):19923-19931. doi:10.1021/acs.jpcc.6b02393
141. Murdoch JR. What is the rate-limiting step of a multistep reaction? *J Chem Educ*. 1981;58(1):32-36. doi:10.1021/ed058p32
  142. Kozuch S, Martin JML. The Rate-Determining Step is Dead. Long Live the Rate-Determining State! *ChemPhysChem*. 2011;12(8):1413-1418. doi:10.1002/cphc.201100137
  143. Laidler KJ. Just what is a transition state? *J Chem Educ*. 1988;65(6):540. doi:10.1021/ed065p540
  144. Sanches-Neto FO, Coutinho ND, Silva V. A novel assessment of the role of the methyl radical and water formation channel in the CH<sub>3</sub>OH + H reaction. *Physical Chemistry Chemical Physics*. Published online 2017. doi:10.1039/C7CP03806B
  145. Coutinho ND, Silva VHC, Mundim KC, de Oliveira HCB. Description of the effect of temperature on food systems using the deformed Arrhenius rate law: deviations from linearity in logarithmic plots vs. inverse temperature. *Rendiconti Lincei*. 2015;26(2):141-149. doi:10.1007/s12210-015-0407-4
  146. Perrin CL. Comment on "Topography of the free energy landscape of Claisen-Schmidt condensation: solvent and temperature effects on the rate-controlling step." *Phys Chem Chem Phys*. 2021;23:22199-22201.
  147. Dantas Coutinho N, Gontijo Machado H, Carvalho-Silva H, Alves da Silva W, Henrique Carvalho V. Reply to "Comment on 'Topography of the free energy landscape of Claisen-Schmidt condensation: solvent and temperature effects on the rate-controlling step.'" *Phys Chem Chem Phys*, 2021, 23, 6738. 2021;23:22202-22206. doi:10.1039/D1CP02799A
  148. Blehrádek J. Influence of Temperature on Biological Processes. *Nature Publishing Group*. 1926;118(2960).
  149. James WO. *Plant Respiration*. Clarendon Press Oxford; 1953.
  150. Carey JC, Tang J, Templer PH, et al. Temperature response of soil respiration largely unaltered with experimental warming. *Proc Natl Acad Sci U S A*. 2016;113(48):13797-13802. doi:10.1073/pnas.1605365113
  151. Wang W, Wang H, Zu Y, Li X, Koike T. Characteristics of the temperature coefficient, Q<sub>10</sub>, for the respiration of non-photosynthetic organs and soils of forest ecosystems. *Front For China*. 2008;7(27):2008-2010. doi:DOI 10.1007/s11461-006-0018-4
  152. Abele D, Heise K, Pörtner HO, Puntarulo S. Temperature-dependence of mitochondrial function and production of reactive oxygen species in the intertidal mud clam *Mya arenaria*. *Journal of Experimental Biology*. 2002;205(13):1831-1841. doi:10.1242/jeb.205.13.1831
  153. Alcaraz M, Almeda R, Saiz E, et al. Effects of temperature on the metabolic stoichiometry of Arctic zooplankton. *Biogeosciences*. 2013;10(2):689-697. doi:10.5194/bg-10-689-2013

154. Barclay CJ. Modelling diffusive O<sub>2</sub> supply to isolated preparations of mammalian skeletal and cardiac muscle. *J Muscle Res Cell Motil.* 2005;26(4-5):225-235. doi:10.1007/s10974-005-9013-x
155. Bennett AF. Thermal dependence of muscle function. *Am J Physiol Regul Integr Comp Physiol.* 1984;16(2). doi:10.1152/ajpregu.1984.247.2.r217
156. Carrasquel-Ursulaez W, Moldenhauer H, Castillo JP, Latorre R, Alvarez O. Biophysical analysis of thermosensitive TRP channels with a special focus on the cold receptor TRPM8. *Temperature.* 2015;2(2):188-200. doi:10.1080/23328940.2015.1047558
157. Clarke A, Fraser KPP. Why does metabolism scale with temperature? BT - Functionnal ecology. 2004;18:243-251.
158. Clarke A. Is there a Universal Temperature Dependence of metabolism? *Funct Ecol.* 2004;18(2):252-256. doi:10.1111/j.0269-8463.2004.00842.x
159. Gillooly JF, Allen AP, Savage VM, Charnov EL, West GB, Brown JH. Response to Clarke and Fraser: effects of temperature on metabolic rate. *Funct Ecol.* 2006;20:400-404.
160. Deban SM, Lappin AK. Thermal effects on the dynamics and motor control of ballistic prey capture in toads: Maintaining high performance at low temperature. *Journal of Experimental Biology.* 2011;214(8):1333-1346. doi:10.1242/jeb.048405
161. Dell AI, Pawar S, Savage VM. The thermal dependence of biological traits. *Ecology.* 2013;94(5):1205-1206. doi:10.1890/12-2060.1
162. Donley JM, Shadwick RE, Sepulveda CA, Syme DA. Thermal dependence of contractile properties of the aerobic locomotor muscle in the leopard shark and shortfin mako shark. *Journal of Experimental Biology.* 2007;210(7):1194-1203. doi:10.1242/jeb.02730
163. Eswaran H, Berg EVANDEN, Reich P. Organic Carbon in Souls of the World. *Soil Science of America Journal.* 1993;57(1):192-194.
164. Held K, Voets T, Vriens J. TRPM3 in temperature sensing and beyond. *Temperature.* 2015;2(2):201-213. doi:10.4161/23328940.2014.988524
165. Kirschbaum MUF. The temperature dependence of organic-matter decomposition - Still a topic of debate. *Soil Biol Biochem.* 2006;38(9):2510-2518. doi:10.1016/j.soilbio.2006.01.030
166. Wood CM, Brix K V., De Boeck G, et al. Mammalian metabolic rates in the hottest fish on earth. *Sci Rep.* 2016;6(May 2016):1-9. doi:10.1038/srep26990
167. Lake SL, MacMillan HA, Williams CM, Sinclair BJ. Static and dynamic approaches yield similar estimates of the thermal sensitivity of insect metabolism. *J Insect Physiol.* 2013;59(8):761-766. doi:10.1016/j.jinsphys.2013.04.010
168. Lloyd J, Taylor JA. On the Temperature Dependence of Soil Respiration. *Funct Ecol.* 1994;8(3):315. doi:10.2307/2389824



169. McCullough JN, Zhang N, Reich DL, et al. Cerebral metabolic suppression during hypothermic circulatory arrest in humans. *Annals of Thoracic Surgery*. 1999;67(6):1895-1899. doi:10.1016/S0003-4975(99)00441-5
170. Chen X, Chen HYH. Global effects of plant litter alterations on soil CO<sub>2</sub> to the atmosphere. *Glob Chang Biol*. 2018;24(8):3462-3471. doi:10.1111/gcb.14147
171. Watanabe YY, Goldman KJ, Caselle JE, Chapman DD, Papastamatiou YP. Comparative analyses of animal-tracking data reveal ecological significance of endothermy in fishes. *Proc Natl Acad Sci U S A*. 2015;112(19):6104-6109. doi:10.1073/pnas.1500316112
172. Huey RB, Kingsolver JG. Variation in universal temperature dependence of biological rates. *Proc Natl Acad Sci U S A*. 2011;108(26):10377-10378. doi:10.1073/pnas.1107430108
173. Marsay CM, Sanders RJ, Henson SA, Pabortsava K, Achterberg EP, Lampitt RS. Attenuation of sinking particulate organic carbon flux through the mesopelagic ocean. *Proc Natl Acad Sci U S A*. 2015;112(4):1089-1094. doi:10.1073/pnas.1415311112
174. King AW, Gunderson CA, Post WM, Weston DJ, Wullschlegel SD. Plant respiration in a warmer world. *Science (1979)*. 2006;312(5773):536-537. doi:10.1126/science.1114166
175. Lin PC, Kreutzer U, Jue T. Anisotropy and temperature dependence of myoglobin translational diffusion in myocardium: Implication for oxygen transport and cellular architecture. *Biophys J*. 2007;92(7):2608-2620. doi:10.1529/biophysj.106.094458
176. Lombard F, Erez J, Michel E, Labeyrie L. Temperature effect on respiration and photosynthesis of the symbiont-bearing planktonic foraminifera *Globigerinoides ruber*, *Orbulina universa*, and *Globigerinella siphonifera*. *Limnol Oceanogr*. 2009;54(1):210-218. doi:10.4319/lo.2009.54.1.0210
177. Streicher JW, Cox CL, Birchard GF. Non-linear scaling of oxygen consumption and heart rate in a very large cockroach species (*Gromphadorhina portentosa*): Correlated changes with body size and temperature. *Journal of Experimental Biology*. 2012;215(7):1137-1143. doi:10.1242/jeb.061143
178. Wythers KR, Reich PB, Bradford JB. Incorporating temperature-sensitive Q<sub>10</sub> and foliar respiration acclimation algorithms modifies modeled ecosystem responses to global change. *J Geophys Res Biogeosci*. 2013;118(1):77-90. doi:10.1029/2011JG001897
179. Xiao Y. Modelling temperature-dependency in biology by generalizing temperature coefficient Q<sub>10</sub>. *Ecol Modell*. 2000;127(2-3):283-290. doi:10.1016/S0304-3800(99)00218-5
180. Heskell MA, O'Sullivan OS, Reich PB, et al. Convergence in the temperature response of leaf respiration across biomes and plant functional types. *Proc Natl Acad Sci U S A*. 2016;113(14):3832-3837. doi:10.1073/pnas.1520282113
181. Stahl WR. Scaling of respiratory variables in mammals. *J Appl Physiol*. 1967;22(3):453-460. doi:10.1152/jappl.1967.22.3.453

182. Buhr ED, Yoo SH, Takahashi JS. Temperature as a universal resetting cue for mammalian circadian oscillators. *Science (1979)*. 2010;330(6002):379-385. doi:10.1126/science.1195262
183. Taylor CR, Weibel ER, Weber JM, et al. Design of the oxygen and substrate pathways. I. Model and strategy to test symmorphosis in a network structure. *Journal of Experimental Biology*. 1996;199(8):1643-1649. doi:10.1242/jeb.199.8.1643
184. Weibel ER, Taylor CR, Weber JM, Vock R, Roberts TJ, Hoppeler H. Design of the oxygen and substrate pathways. VII. Different structural limits for oxygen and substrate supply to muscle mitochondria. *Journal of Experimental Biology*. 1996;199(8):1699-1709. doi:10.1242/jeb.199.8.1699
185. HOLLAND WC, KLEIN RL. Effects of temperature, Na and K concentration and quinidine on transmembrane flux of K<sup>42</sup> and incidence of atrial fibrillation. *Circ Res*. 1958;6(4):516-521. doi:10.1161/01.RES.6.4.516
186. Park DS, Shekhar A, Marra C, et al. Fhf2 gene deletion causes temperature-sensitive cardiac conduction failure. *Nat Commun*. 2016;7(May):1-10. doi:10.1038/ncomms12966
187. Yang K, He R, Yang W, et al. Temperature response of soil carbon decomposition depends strongly on forest management practice and soil layer on the eastern Tibetan Plateau. *Sci Rep*. 2017;7(1):1-8. doi:10.1038/s41598-017-05141-2
188. Tang J, Cheng H, Fang C. The temperature sensitivity of soil organic carbon decomposition is not related to labile and recalcitrant carbon. *PLoS One*. 2017;12(11):1-15. doi:10.1371/journal.pone.0186675
189. Yan D, Li J, Pei J, Cui J, Nie M, Fang C. The temperature sensitivity of soil organic carbon decomposition is greater in subsoil than in topsoil during laboratory incubation. *Sci Rep*. 2017;7(1):1-9. doi:10.1038/s41598-017-05293-1
190. Liu Z, Ballantyne AP, Poulter B, et al. Precipitation thresholds regulate net carbon exchange at the continental scale. *Nat Commun*. 2018;9(1):1-10. doi:10.1038/s41467-018-05948-1
191. Ito E, Ikemoto Y, Yoshioka T. Thermodynamic implications of high Q<sub>10</sub> of thermo-TRP channels in living cells. *Biophysics (Japan)*. 2015;11(0):33-38. doi:10.2142/biophysics.11.33
192. Laursen WJ, Anderson EO, Hoffstaetter LJ, Bagriantsev SN, Gracheva EO. Species-specific temperature sensitivity of TRPA1. *Temperature*. 2015;2(2):214-226. doi:10.1080/23328940.2014.1000702
193. Raddatz N, Castillo JP, Gonzalez C, Alvarez O, Latorre R. Temperature and voltage coupling to channel opening in transient receptor potential melastatin 8 (TRPM8). *Journal of Biological Chemistry*. 2014;289(51):35438-35454. doi:10.1074/jbc.M114.612713

194. Chang RL, Andrews K, Kim D, Li Z, Godzik A, Palsson BO. Structural systems biology evaluation of metabolic thermotolerance in *Escherichia coli*. *Science* (1979). 2013;340(6137):1220-1223. doi:10.1126/science.1234012
195. Halsey LG, Matthews PGD, Rezende EL, Chauvaud L, Robson AA. The interactions between temperature and activity levels in driving metabolic rate: theory, with empirical validation from contrasting ectotherms. *Oecologia*. 2015;177(4):1117-1129. doi:10.1007/s00442-014-3190-5
196. Reyes BA, Pendergast JS, Yamazaki S. Mammalian peripheral circadian oscillators are temperature compensated. *J Biol Rhythms*. 2008;23(1):95-98. doi:10.1177/0748730407311855
197. Schulte PM. The effects of temperature on aerobic metabolism: Towards a mechanistic understanding of the responses of ectotherms to a changing environment. *Journal of Experimental Biology*. 2015;218(12):1856-1866. doi:10.1242/jeb.118851
198. Stanley SM. Relation of phanerozoic stable isotope excursions to climate, bacterial metabolism, and major extinctions. *Proc Natl Acad Sci U S A*. 2010;107(45):19185-19189. doi:10.1073/pnas.1012833107
199. Hararuk O, Smith MJ, Luo Y. Microbial models with data-driven parameters predict stronger soil carbon responses to climate change. *Glob Chang Biol*. 2015;21(6):2439-2453. doi:10.1111/gcb.12827
200. Lal R. Soil carbon sequestration impacts on global climate change and food security. *Science* (1979). 2004;304(5677):1623-1627. doi:10.1126/science.1097396
201. Mukherjee J, Ray S, Ghosh PB. A system dynamic modeling of carbon cycle from mangrove litter to the adjacent Hooghly estuary, India. *Ecol Modell*. 2013;252(1):185-195. doi:10.1016/j.ecolmodel.2012.06.036
202. Wang WJ, Wang HM, Zu YG, Li XY, Koike T. Characteristics of Root, Stem and Soil Respiration Q10 Temperature Coefficients in Forest Ecosystems. *Acta Phytocologica Sinica*. 2005;29(4):680-691.
203. Schulte PM, Wythers KR, Reich PPB, et al. Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Journal of Experimental Biology*. 2015;7(2):201-213. doi:10.1242/jeb.061143
204. Fangué NA, Richards JG, Schulte PM. Do mitochondrial properties explain intraspecific variation in thermal tolerance? *Journal of Experimental Biology*. 2009;212(4):514-522. doi:10.1242/jeb.024034
205. Li J, Hibbert DB, Fuller S, Vaughn G. A comparative study of point-to-point algorithms for matching spectra. *Chemometrics and Intelligent Laboratory Systems*. 2006;82(1-2 SPEC. ISS):50-58. doi:10.1016/j.chemolab.2005.05.015
206. Linstrom PJ, Mallard WG. The NIST Chemistry WebBook: A Chemical Data Resource on the Internet. *J Chem Eng Data*. 2001;46(5).
207. McCulloch WS, Pitts W. A Logical Calculus of the Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics*. 1943;5:115-133. doi:10.1007/bf02478259

208. Rosenblatt F. The Perceptron - A Perceiving and Recognizing Automaton. *Report 85, Cornell Aeronautical Laboratory*. Published online 1957:460-461.
209. Minsky M. A Neural-Analogue Calculator Based upon a Probability Model of Reinforcement. *Harvard University Psychological Laboratories*,. Published online 1952.
210. Widrow B. An Adaptive "Adaline" Neuron Using Chemical "Memistors." *Stanford Electronics Laboratories Technical Report*. Published online 1960:1553-2.
211. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature*. 1986;323(6088):533-536. doi:10.1038/323533a0
212. LeCun Y, Boser B, Denker JS, et al. Handwritten Digit Recognition with a Back-Propagation Network. *AT&T Bell Laboratories*. 1989;(07733):396-404.
213. Weng J, Ahuja N, Huang TS. Cresceptron: A Self-Organizing Neural Network Which Grows Adaptively. *RN*. 1992;63(2):576-581. doi:10.1109/IJCNN.1992.287150
214. Hearst MartiA, Scholkopf Bernhard, Dumais Susan, Osuna Edgar, Platt J. Support vector machines. *IEEE Intelligent Systems and their Applications*. 1998;13(4):18-28.
215. Hochreiter S, Schmidhuber J. Long Short-Term Memory. *Neural Comput*. 1997;(9):1735–1780. doi:10.1162/neco.1997.9.8.1735
216. Graves A, Mohamed A rahman, Hinton G. Speech Recognition With Deep Recurrent Neural Networks. *IEEE*. 2013;(3). doi:https://doi.org/10.1109/ICASSP.2013.6638947
217. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-Based Learning Applied to Document Recognition. *proc OF THE IEEE*. Published online 1998.
218. Fei-Fei L, Deng J, Li K. ImageNet: A Large-Scale Hierarchical Image Database. *Journal of Vision - IEEE*. 2009;9(8):1037-1037. doi:https://doi.org/10.1109/CVPR.2009.5206848
219. Russakovsky O, Deng J, Su H, et al. ImageNet Large Scale Visual Recognition Challenge. Published online 2015.
220. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolution Neural Networks. *Adv Neural Inf Process Syst*. 2012;60(6):84-90. doi:http://dx.doi.org/10.1145/3065386
221. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. Published online 2015:1-14.
222. Targ S, Almeida D, Lyman K. Resnet in Resnet: Generalizing Residual Architectures. Published online 2016:1-7.
223. Canziani A, Culurciello E, Paszke A. An Analysis of Deep Neural Network Models for Practical Applications. Published online 2017:1-7.
224. Ramsundar B, Eastman P, Walters P, Pande V. *Deep Learning for the Life Science*. 1st ed. (Tache N, Loukides M, Tozer K, Head R, eds.). O'Reilly Media; 2019.

225. Kingma DP, Ba JL. Adam: A method for stochastic optimization. In: *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. ; 2015:1-15.
226. Mannor S, Peleg D, Rubinstein R. The Cross Entropy Method for Classification. In: *Proceedings of the 22nd International Conference on Machine Learning*. Association for Computing Machinery; 2005:561–568.
227. Goutte C, Gaussier E. Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. In: *Proceedings of the 27th European Conference on Advances in Information Retrieval Research*. Springer-Verlag; 2005:345–359.
228. Mcgregor MJ, Pallai P v. Clustering of Large Databases of Compounds: Using the MDL “Keys” as Structural Descriptors. *J Chem Inf Comput Sci*. 1997;37(3):443-448. doi:<https://doi.org/10.1021/ci960151e>
229. Rogers D, Hahn M. Extended-connectivity fingerprints. *J Chem Inf Model*. 2010;50(5):742-754. doi:10.1021/ci100050t
230. Landrum G. RDKit: Open-Source Cheminformatics Software. Published online 2016.
231. Howard J and others. fastai. Published online 2018.
232. Sato S. Tunneling in bimolecular reactions. *Chem Phys*. 2005;315(1-2):65-75. doi:10.1016/j.chemphys.2005.01.034
233. CPMD, version 4.1. Published online 2012.
234. Hallett J. The temperature dependence of the viscosity of supercooled water. *Proceedings of the Physical Society*. 1963;82(6):1046-1050. doi:10.1088/0370-1328/82/6/326
235. Savitzky Abraham, Golay MJE. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal Chem*. 1964;36(8):1627-1639. doi:<https://doi.org/10.1021/ac60214a047>



## APÊNDICE A – PUBLICAÇÕES

### A.1 – “Transitivity”: A code for Computing Kinetic and Related Parameters in Chemical Transformations and Transport Phenomena



Article

#### “Transitivity”: A Code for Computing Kinetic and Related Parameters in Chemical Transformations and Transport Phenomena

Hugo G. Machado <sup>1,2,\*</sup>, Flávio O. Sanches-Neto <sup>1,2,\*</sup>, Nayara D. Coutinho <sup>3,\*</sup>, Kleber C. Mundim <sup>2</sup> , Federico Palazzetti <sup>3,\*</sup> and Valter H. Carvalho-Silva <sup>1,2,\*</sup> 

<sup>1</sup> Grupo de Química Teórica e Estrutural de Anápolis, Centro de Pesquisa e Pós-Graduação, Universidade Estadual de Goiás, 75132-400 Anápolis, GO, Brazil

<sup>2</sup> Instituto de Química, Universidade de Brasília, Caixa Postal 4478, 70904-970 Brasília, Brazil; kcmundim@unb.br

<sup>3</sup> Dipartimento di Chimica, Biologia e Biotecnologie, Università di Perugia, 06123 Perugia, Italy

\* Correspondence: hugogontijomachado@gmail.com (H.G.M.); flavio\_olimpio@outlook.com (F.O.S.-N.); nayaradcoutinho@gmail.com (N.D.C.); fede\_75it@yahoo.it (F.P.); fatioleg@gmail.com (V.H.C.-S.); Tel.: +55-62-3328-1156 (V.H.C.-S)

Academic Editor: Pier Luigi Gentili

Received: 16 August 2019; Accepted: 8 September 2019; Published: 25 September 2019



**Abstract:** The Transitivity function, defined in terms of the reciprocal of the apparent activation energy, measures the propensity for a reaction to proceed and can provide a tool for implementing phenomenological kinetic models. Applications to systems which deviate from the Arrhenius law at low temperature encouraged the development of a user-friendly graphical interface for estimating the kinetic and thermodynamic parameters of physical and chemical processes. Here, we document the Transitivity code, written in Python, a free open-source code compatible with Windows, Linux and macOS platforms. Procedures are made available to evaluate the phenomenology of the temperature dependence of rate constants for processes from the Arrhenius and Transitivity plots. Reaction rate constants can be calculated by the traditional Transition-State Theory using a set of one-dimensional tunneling corrections (Bell (1935), Bell (1958), Skodje and Truhlar and, in particular, the deformed (*d*-TST) approach). To account for the solvent effect on reaction rate constant, implementation is given of the Kramers and of Collins–Kimball formulations. An input file generator is provided to run various molecular dynamics approaches in CPMD code. Examples are worked out and made available for testing. The novelty of this code is its general scope and particular exploit of *d*-formulations to cope with non-Arrhenius behavior at low temperatures, a topic which is the focus of recent intense investigations. We expect that this code serves as a quick and practical tool for data documentation from electronic structure calculations: It presents a very intuitive graphical interface which we believe to provide an excellent working tool for researchers and as courseware to teach statistical thermodynamics, thermochemistry, kinetics, and related areas.

**Keywords:** *d*-TST; activation energy; Transitivity plot; solution kinetic

#### 1. Introduction

Recent applications of chemical kinetics to a variety of complex systems involves the accurate dealing of properties to be described by techniques, which treat a series of processes beyond elementary chemical quantum dynamics or even approximate classical and semiclassical approaches. We analyze in a companion paper [1] the state of our approaches to these problems from a general viewpoint. Here, we deal with explicit computational calculations that allow moving directly in a simple way to global applications.



## A.2 – Temperature coefficient ( $Q_{10}$ ) and its applications in biological systems: Beyond the Arrhenius theory

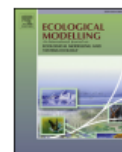
Ecological Modelling 431 (2020) 109127



Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: [www.elsevier.com/locate/ecolmodel](http://www.elsevier.com/locate/ecolmodel)



Review

### Temperature coefficient ( $Q_{10}$ ) and its applications in biological systems: Beyond the Arrhenius theory



Kleber C. Mundim<sup>a,\*</sup>, Solange Baraldi<sup>b</sup>, Hugo G. Machado<sup>a</sup>, Fernando M.C. Vieira<sup>a</sup>

<sup>a</sup> Institute of Chemistry, Brasília University, P.O. Box 4478, 70904-970, Brasília, Brazil

<sup>b</sup> Faculty of Health Sciences-Nursing Department, University of Brasília, P.O. Box 4478, 70904-970, Brasília, Brazil

#### ARTICLE INFO

##### Keywords:

$Q_{10}$  temperature sensitivity  
global climate models  
deformed Arrhenius theory  
soil organic carbon (SOC)  
rates on biological processes  
metabolism of plants and animals

#### ABSTRACT

The  $Q_{10}$  temperature coefficient, which is widely used in scientific literature, is a measure of the temperature sensitivity of chemical reaction rates or biological processes. However, the conclusions drawn from applying this coefficient to experimental data obtained from biological processes are not universal. In many biological processes,  $Q_{10}$  values are often discordant with the results predicted by the Arrhenius law. The hypothesis tested in the present study is that this problem arises mainly from the fact that the  $Q_{10}$  coefficient is defined by the ratio between rates described by exponential laws instead of power laws. Considering this hypothesis and the need to review the mathematical laws and models currently used to describe rates and  $Q_{10}$  coefficients, we propose a model beyond the usual Arrhenius theory or exponential decay law herein. The proposed mathematical model is based on the theory of deformed exponential functions, with the ordinary  $Q_{10}$  model representing the conventional exponential function. Therefore, all results following the standard model remain valid. Moreover, we include a  $Q_{10}$  free open-source code, written in Python, and compatible with Windows, Linux and macOS platforms. The validation of the proposed model and confirmation of the given hypothesis were performed based on the following temperature-dependent biological processes: soil organic carbon (SOC) decomposition (which is essential to forecast the impact of climate change on terrestrial ecosystems); the metabolism of Arctic zooplankton; physiological processes of the respiratory and cardiovascular systems; rate of oxygen consumption in mitochondria of the eurythermal killifish *Fundulus heteroclitus*, and leaf respiration.

#### 1. Introduction

The  $Q_{10}$  coefficient (James, 1953; Blehrádek; Hegarty) has been widely described in the literature as a measure of the degree of temperature dependence of a given chemical or biological process. It is usually defined as the rate ratio of a given process taking place at temperatures differing by 10 units ( $^{\circ}\text{C}$  or  $\text{K}$ ). However, conclusions drawn by different researchers from the use of this coefficient are not universal (Wang et al., 2006; Carey) and are not always consistent with each other (Abele et al., 2002; Alcaraz et al., 2012; Barclay, 2005; Bennett, 1984; Carrasquel-Ursulaez et al., 2015; Chang et al., 2013; Clark et al., 2013; Clarke, 2004; Clarke and Fraser, 2004; Deban and Lappin, 2011; Dell et al., 2013; Ding et al., 2018; Donley et al., 2007; Eswaran et al., 1993; Fangue et al., 2009; Gillooly et al., 2006; Halsey et al., 2015; Held et al., 2015; Helmuth et al., 2004; Holland and Klein, 1958; Ito et al., 2015; King et al., 2006; Kirschbaum, 2006; Lake et al., 2013; Lal, 2004; Laursen et al., 2015; Lin et al., 2007; Liu et al., 2018; Lloyd and Taylor, 1994; Lombard et al., 2009; McCullough et al.,

1999; Raddatz et al., 2014; Reyes et al., 2008; Schulte, 2015; Stanley, 2010; Streicher et al., 2012; Tang et al., 2017; Taylor et al., 1996; Wang et al., 2005; Weibel et al., 1996; Wythers et al., 2013; Xiao, 2000; Yan et al., 2017; Yang et al., 2017; Zhang et al., 2015; Park et al.,; Davidson et al.,; Kramshøj et al.,; Wood et al.,). Therefore, it is reasonable to presume that this problem arises from the fact that the  $Q_{10}$  coefficient is defined by the ratio between rates described by exponential laws instead of power laws, which leads to deviations from linearity in the Arrhenius plots. This hypothesis may be tested in experiments conducted by several research groups (Abdelsayed et al., 2018; Barton and Yvon-Durocher, 2019; Han and Jin, 2018; Hoyt et al., 2019; Magnabosco et al., 2018; Makhnykina et al., 2018; Wudarski et al., 2019). In these cases, a temperature increase by 10 units does not entail the doubling or tripling of the rate of a given process, as is usually assumed in the relevant literature (Heskel et al., 2016).

Based on this hypothesis, an alternative model is proposed that generalises the  $Q_{10}$  coefficient and unifies the conflicting results obtained by researchers. Overall, the primary objective of the present

\* Corresponding author: Kleber Carlos Mundim Institute of Chemistry, Brasília University, P.O. Box 4478, 70904-970, Brasília, Brazil  
E-mail address: [kcmundim@unb.br](mailto:kcmundim@unb.br) (K.C. Mundim).

<https://doi.org/10.1016/j.ecolmodel.2020.109127>

Received 21 December 2019; Received in revised form 25 April 2020; Accepted 9 May 2020  
0304-3800/ © 2020 Elsevier B.V. All rights reserved.

## A.3 – Topography of the free energy landscape of Claisen-Schmidt condensation: solvent and temperature effects on the rate-controlling step

PCCP



PAPER

### Topography of the free energy landscape of Claisen–Schmidt condensation: solvent and temperature effects on the rate-controlling step†

Cite this: DOI: 10.1039/d0cp05659f

Nayara Dantas Coutinho,<sup>a</sup> Hugo Gontijo Machado,<sup>b</sup> Valter Henrique Carvalho-Silva<sup>b</sup> and Wender Alves da Silva<sup>a</sup>

Received 29th October 2020,  
Accepted 4th January 2021

DOI: 10.1039/d0cp05659f

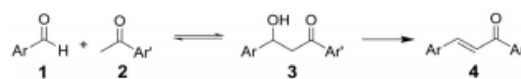
rsc.li/pccp

Recent studies have found that hydroxide elimination and the C=C bond formation step in base-promoted aldol condensation have a strong influence on the overall rate of the reaction, in contrast to the well-accepted first enolization or C–C bond formation step. Here, applying theoretical models to the prototypical reaction of chalcone formation, the complete free energy profile of Claisen–Schmidt condensation is assessed, revealing how a protic solvent and a slight increase in temperature can induce the second enolization as the rate-controlling step. It is also shown that the overall inverse kinetic isotope effect does not exclude the second enolization as a rate-controlling step in protic continuum medium.

### Introduction

The Claisen–Schmidt reaction is a well-known protocol for the synthesis of chalcone **4**<sup>1,2</sup> and can be efficiently achieved by the base-catalyzed condensation of benzaldehyde **1** with acetophenone **2**, a classic representative case of aldol condensation<sup>3,4</sup> for formation of chalcone **4**.<sup>5–8</sup>

Scheme 1 shows the widely accepted key steps for the Claisen–Schmidt condensation mechanism. Multi-step reactions can be limited by kinetic and thermodynamic aspects by controlling the general velocity constant in aldolic condensation. In the last eight decades, several studies have allowed the elucidation of the main steps of its mechanism, Fig. 1: (i) the first study trying to explain these effects was first reported in 1937,<sup>9–11</sup> for the condensation of an aldol in concentrated solutions of non-aromatic aldehydes, as a first order reaction, and assuming the removal of protons and the formation of enolate **2**<sup>–</sup> (step **R1**) as the rate controlling step (RCS) – recent work concluded that this step has a strong effect on the overall



Scheme 1 General formation of a chalcone via ketol and E1cb elimination.

<sup>a</sup> Laboratory of Bioactive Compounds Synthesis N.T.S., University of Brasilia (IQ-UnB), Campus Universitário Darcy Ribeiro, Brasilia, DF, Brazil.

E-mail: nayaradcoutinho@gmail.com, wender@unb.br

<sup>b</sup> Modeling of Physical and Chemical Transformations Division, Theoretical and Structural Chemistry Group, Research and Graduate Center, Goiás State University, 75132-903, Anapolis, Brazil. E-mail: fatioleg@gmail.com

† Electronic supplementary information (ESI) available: Details of the Kramers formulations, reaction rate constant for each step of chalcone formation, fitted equations of the temperature dependence of the elementary rate constants, and cartesian coordinates of all stationary points (reactants, intermediates, transition states and products). See DOI: 10.1039/d0cp05659f

rate;<sup>12–14</sup> (ii) however, Coombs and Evans<sup>11</sup> proved that the reaction rate law for chalcone formation follows  $v = k_{\text{obs}}[1][2][\text{Base Catalyst}]$  with C–C bond formation as the slowest step; and (iii) analogous findings were obtained in other studies,<sup>15,16</sup> consolidating step **R2** as the rate-controlling step, which influenced later papers on aldol condensation mechanisms.<sup>6,17,18</sup> In 2016, Chang and Perrin<sup>19</sup> established a complete mechanism for aldol condensation, and experimentally proved **R5**, hydroxide elimination and C=C bond formation, to be a rate-controlling step, leading to a revision of the previous understanding. Details about the elementary steps in chalcone formation are shown in Scheme S1 (ESI†) and in Fig. 1.

Chang and Perrin's conclusion was experimentally supported by: (i) the faster reversion from ketol **3** to the reactants **1** and **2** when compared to the progression to **4**; (ii) the solvent isotope effect, which demonstrates faster chalcone formation from reactants **1** and **2** in D<sub>2</sub>O when compared to H<sub>2</sub>O, eliminating the second enolization, **R4**, as the rate-controlling step; and (iii) the construction of the Gibbs free energy diagram from the combination of their experimental results and others. However, experimental constraints addressed by Chang and Perrin<sup>19</sup> can merge solvent and isotope effects, making it hard to assess these effects



**A.4 – Reply to the ‘Comment on “Topography of the Free Energy Landscape on the Claisen–Schmidt Condensation: Solvent and Temperature Effect in the Rate-Controlling Step” by N. D. Coutinho, H. G. Machado, V. H. Carvalho-Silva and W. A. da Silva, *Phys. Chem. Chem. Phys.*, 2021, 23, 6738’**



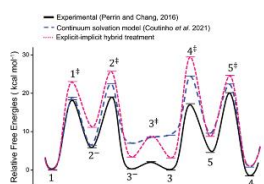
PCCP

COMMENT

Cite this: DOI: 10.1039/d1cp02799a

**Reply to the ‘Comment on “Topography of the Free Energy Landscape on the Claisen–Schmidt Condensation: Solvent and Temperature Effect in the Rate-Controlling Step” by N. D. Coutinho, H. G. Machado, V. H. Carvalho-Silva and W. A. da Silva, *Phys. Chem. Chem. Phys.*, 2021, 23, 6738†**

Nayara Dantas Coutinho,<sup>a</sup> Hugo Gontijo Machado,<sup>b</sup> Valter Henrique Carvalho-Silva<sup>b</sup> and Wender Alves da Silva<sup>a</sup>



Received 21st June 2021,  
Accepted 13th August 2021

DOI: 10.1039/d1cp02799a

rsc.li/pccp

## Introduction

Perrin has commented on our recent theoretical studies on the Claisen–Schmidt Condensation<sup>1</sup> claiming that our calculations were flawed. However, herein we show several inconsistencies in the Comment, and each of these issues addressed in his Comment will be discussed in turn. Before further discussion, we will emphasize again the main conclusions of our work:

(i) The nonexistence of a step with a much higher energetic barrier than the others, in both water and acetonitrile

<sup>a</sup> Laboratory of Bioactive Compounds Synthesis N.T.S., University of Brasilia (IQ-UnB), Campus Universitário Darcy Ribeiro, Brasilia, DF, Brazil. E-mail: nayaradcoutinho@gmail.com, wender@unb.br

<sup>b</sup> Modeling of Physical and Chemical Transformations Division, Theoretical and Structural Chemistry Group, Research and Graduate Center, Góias State University, 75132-903, Anápolis, Brazil. E-mail: fatioleg@gmail.com

† Electronic supplementary information (ESI) available: Cartesian coordinate of all stationary points with explicit water molecules. See DOI: 10.1039/d1cp02799a

In the Comment on our paper on the description of the Gibbs Free energy profile of Claisen–Schmidt condensation, it is claimed that our calculations are flawed due to inconsistencies with experimental isotope effects in aqueous acetonitrile. In this Reply, we presented rigorous arguments, ambiguities in the Comment and new calculations confirming the consistency of our results: (i) small differences in the relative energetic barriers in both experimental and theoretical curves make the assignment of the rate-limiting step debatable, making the concept of RCS questionable; (ii) it is shown how the misinterpretation of the elementary steps and of the overall processes rate constants led the Comment to incorrect conclusions about the behavior of the inverse isotopic effect; (iii) neglect in the Comment of the inverse kinetic isotope effect in step **R2** due to the hybridization conversion, and of the inverse equilibrium isotope effect for step **R1** to describe an overall iKIE > 1, (iv) an erroneous suggestion in the Comment that the disagreement between experimental kinetic parameters is due to the fact that acetonitrile is not used in previous experimental works, when contradictorily the literature recommends it as being indispensable to allow kinetic accuracy; and (v) new calculations improved by explicit-implicit hybrid treatment again ensure that step **R4**, and not step **R5**, can assume the role of RCS in protic solvents. Recognizing that questioning is an excellent path for promoting understanding, we hope that the answers provided here will help to clarify and expand the pertinent topics under discussion.

continuum medium, making the concept of Rate-Controlling Step (RCS) debatable.

(ii) The relative free energy estimated experimentally by ref. 2 and calculated by us showed that the difference between the elementary energy barriers of steps **R2** and **R5** do not exceed 1 kcal mol<sup>-1</sup>;

(iii) Our calculations corroborated Perrin and Chang's conclusion supporting the faster reversion from ketol 3 to reactants 1 and 2 when compared with progression to 4 in both water and acetonitrile continuum medium;

(iv) Elementary step **R4** showed iKIE < 1 in consonance with assessment in the Comment.

(v) Elementary steps **R2** and **R5** showed iKIE > 1; as a corollary, our calculations showed that an overall iKIE > 1 does not eliminate step **R4** as RCS, since step **R2** supplies this behavior.

We should also highlight here that among all the 5 conclusions underscored above, only item (v) does not agree with the

## APÊNDICE B – LIDANDO COM AS INTERFACES GRÁFICAS DOS CÓDIGOS E MODELOS DESENVOLVIDOS

### B.1 - Download, Instalação e Execução dos Softwares

Os códigos *Transitivity* e dQ<sub>10</sub>-GSA escritos na linguagem Python3 são livres, gratuitos e compatíveis com os sistemas operacionais Windows, Linux e MacOS. Ambos podem ser encontrados em diretórios contidos no endereço eletrônico do GitHub do autor: [github.com/hugogontijomachado](https://github.com/hugogontijomachado). O código *Transitivity* também pode ser baixado através do endereço eletrônico de um dos grupos de pesquisa ao qual o autor faz parte: [mphyschem.com/transitivity](https://mphyschem.com/transitivity). Neste último link o usuário também tem acesso à um manual do programa, arquivos para exemplos de aplicação e até mesmo a um vídeo ilustrativo que auxilia na instalação do programa. No GitHub do autor também é possível ter acesso aos modelos de deep learning aqui desenvolvidos, bem como o acesso a aplicação web “*SpectralAI*” para utilização simples e prática dos modelos.

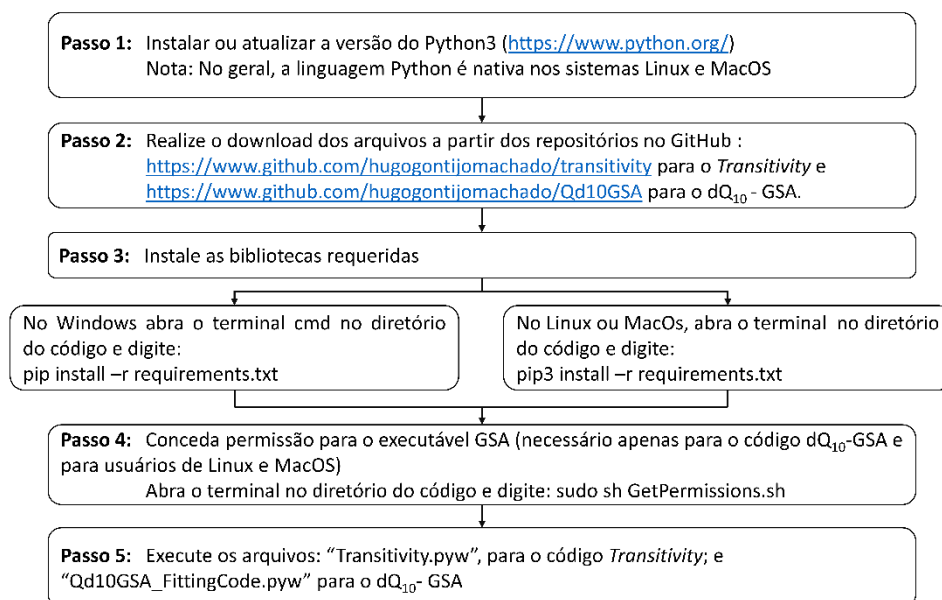
A linguagem Python é nativa dos sistemas Linux e MacOS, mas deve ser instalada manualmente por usuários do sistema operacional Windows. É importante observar que a linguagem requerida para funcionamento dos códigos é o Python3, portanto, usuários de sistemas que possuem nativamente a versão Python2 instalada devem atualizar suas versões. Além disso, diversas bibliotecas Python foram utilizadas no desenvolvimento do código e algumas devem ser instaladas manualmente qualquer que seja o sistema operacional do usuário, estas bibliotecas estão listadas nos arquivos ‘*requirements.txt*’ contidos nos repositórios de cada um dos códigos. Para correta instalação e execução dos códigos siga o passo a passo descrito na Figura 29.

Pesquisadores que utilizarem os *softwares Transitivity* e dQ<sub>10</sub>-GSA em suas pesquisas ou para preparação de suas publicações, podem citá-los através das referências<sup>31</sup> e<sup>32</sup>.

Caso o usuário enfrente dificuldades na utilização dos *softwares* é possível acessar um menu de ajuda nas janelas principais de cada interface.

Se forem observados comportamentos não esperados que possam ter sido causados por *bugs* ou mau funcionamento do código, o autor pede para que os usuários descrevam o problema em detalhes, anexem os resultados (arquivos de

saída, gráficos ou outros arquivos relacionados) e reportem para o e-mail [hugogontijomachado@gmail.com](mailto:hugogontijomachado@gmail.com). Dúvidas e sugestões também são bem-vindas a fim de contribuir com o desenvolvimento e performance dos *softwares*.



**Figura 29:** Passo a passo da instalação e configuração dos programas Transitivity e dQ<sub>10</sub>-GSA.

Adicionalmente, estão em desenvolvimento aplicações *web* ainda mais práticas e intuitivas que irão substituir as interfaces aqui apresentadas, assim como a aplicação *web* para modelos de *deep learning* para análise espectral que pode ser encontrada no repositório *github* do autor: <https://www.github.com/hugogontijomachado>. Deste modo, não será necessário realizar nenhum dos passos de instalação e configuração, o usuário irá apenas acessar um *link* no seu navegador e poderá utilizar as funcionalidades dos algoritmos propostos.

## B.2 - Lidando com o código Transitivity

Em suma, o código *Transitivity* pode estimar parâmetros fenomenológicos a partir dos *plots* de Arrhenius e da transitividade, utilizando o GSA<sup>7</sup>, com diversas opções: Arrhenius, AM<sup>30</sup>, VFT<sup>45</sup>, NTS<sup>47,232</sup> e ASCC<sup>27</sup>. Além disso, o código também permite estimar constantes cinéticas para reações uni e bimoleculares utilizando a TST tradicional com correções de Bell<sup>35</sup><sup>70</sup>, Bell<sup>58</sup><sup>72</sup>, Skodje-Truhlar (ST)<sup>75</sup> e a *d*-

TST<sup>61</sup>. Efeitos de solvente podem ser aplicados pelos modelos de *Collins-Kimball*<sup>66</sup> e *Kramer*<sup>78</sup>.

Outra opção disponível é a criação de *inputs* para diversas dinâmicas moleculares – dinâmica de Born-Oppenheimer (BOMD, do inglês: *Born-Oppenheimer Molecular Dynamics*), Car-Parrinello (CPMD, do inglês: *Car-Parrinello Molecular Dynamics*), meta dinâmica (MTD), Integrais de Trajetória (PIMD, do inglês: *Path-Integral Molecular Dynamics*) e Salto de Superfície de Trajetória (TSH, do inglês: *Trajectory Surface Hopping*) – para utilização do código CMPD<sup>233</sup>.

### B.2.1 - Cinética e Parâmetros Relacionados

Se a opção escolhida for “*Kinetic and Related Parameters*”, o programa precisará dos *outputs* (arquivos de saída) de cálculos de estrutura eletrônica fornecidos pelo programa *Gaussian*<sup>102</sup> dos reagentes, estado de transição e dos produtos. Com isso, o programa calcula e fornece as constantes cinéticas de reação para reações uni e bimoleculares.

Em uma nova janela o usuário pode escolher entre as teorias TST ou *d*-TST<sup>61</sup> e entre as correções de tunelamento e *Bell35*<sup>70</sup>, *Bell58*<sup>72</sup>, *Skodje – Truhlar (ST)*<sup>75</sup>.

Além disso o programa fornece uma visualização do gráfico de *Arrhenius* com a possibilidade de incluir dados experimentais ou teóricos para efeito de comparação e validação. Na mesma janela, propriedades como energia, entalpia, energia livre de *Gibbs*, altura da barreira, parâmetro *d*, frequência imaginária ( $\nu^\ddagger$ ), temperatura de *crossover* ( $T_c$ ) e os parâmetros do modelo de *Skodje* e *Truhlar* podem ser visualizados, além de um gráfico da superfície de energia potencial da reação.

Se os cálculos de estrutura eletrônica de todas as entidades foram realizados utilizando a *keyword* VOLUME, ficará disponível também uma nova janela onde serão fornecidas as constantes cinéticas da reação considerando os efeitos de solvente de acordo com as formulações de *Kramer* e *Collins-Kimball*. O usuário pode fornecer os parâmetros do solvente para estimativa da viscosidade clicando em “*Solvent Type*”. Caso não seja fornecido nenhum parâmetro, por padrão o programa realizará os cálculos utilizando a água como solvente. A viscosidade da água em função da temperatura, utilizando dados experimentais<sup>234</sup>, pode ser expressa como:  $\eta(T) = 2,7024 \cdot 10^{-4} \text{Poise}(1 - 213,0543/T)^{-2,75634}$ . Se a formulação de *Kramer* for selecionada, o programa fornece: o coeficiente de fricção do solvente ( $\mu/s^{-1}$ ), o

coeficiente de transmissão de *Kramer* e a constante global da reação ( $k_{obs}/cm^3 \cdot mol^{-1} \cdot s^{-1}$  ou  $s^{-1}$ ). Se a formulação de *Collins-Kimball* for selecionada, o programa fornece: a constante cinética global da reação ( $k_{obs}/cm^3 \cdot mol^{-1} \cdot s^{-1}$ ) e os coeficientes de difusão ( $cm^2 \cdot s^{-1}$ ) separadamente.

### B.2.2 - Ajuste de Otimização GSA

Se a opção escolhida for “*GSA Fitting*”, os valores da constante cinética e das temperaturas em que foram realizadas as medições serão necessários. O que o programa faz é realizar um processo otimização estocástico utilizando o algoritmo GSA (ver Seção 1.3), para isso, utiliza os dados fornecidos acerca da constante de reação para realizar ajustes de função para as formulações de *Arrhenius*, *Aquilanti-Mundim*, *VFT*, *NTS* e *ASCC*. Para facilitar o ajuste, é possível ao usuário fornecer parâmetros iniciais das formulações:  $d$ ,  $\varepsilon^\ddagger$ ,  $E_a$ ,  $E_0$ ,  $E_v$ ,  $B$  e  $T_0$  (ver formulações da Tabela 1). As informações detalhadas sobre o ajuste ficarão disponíveis no arquivo de saída “Fit\_[FormulaName].dat”. Parâmetros internos do algoritmo de otimização GSA também podem ser controlados<sup>7</sup>. O ajuste no gráfico de transitividade é implementado para a constante cinética de reação apenas utilizando as formulações de *Arrhenius*, *Aquilanti-Mundim* e *VFT*. O gráfico da transitividade é calculado utilizando diferenciação numérica com a opção “*Preview*” e se necessário a suavização dos dados pode ser realizada selecionando a opção “*Apply SG*” habilitando assim o filtro de *Savitzky-Golay*<sup>235</sup>. O programa permite visualizar dinamicamente o passo a passo da otimização nos gráficos da transitividade e de *Arrhenius* ao selecionar a opção “*Animation*”.

### B.2.3 - Dinâmica Molecular

Adicionalmente no *software Transitivity* foi implementada a “*Molecular Dynamic*” onde é possível criar *inputs* de cálculos de dinâmica molecular para o programa CPMD. Para isso, é necessário selecionar o arquivo do sistema a ser estudado, contendo a geometria molecular do sistema em alguma das extensões específicas (\*.xyz, \*.gjf, \*.out e \*.log) e selecionar o método pretendido (BOMD, CPMD, MTD, PIMD ou TSH). Além disso o usuário também tem a opção de escolher o funcional relacionado a Teoria do Funcional da Densidade (DFT, do inglês: *Density*

*Functional Theory*), o pseudopotencial, a temperatura, a carga, o tempo de simulação e o tempo de cada passo. O tamanho da caixa que irá conter o sistema pode ser alterado na seção “*Lattices*”. Adicionalmente, ao criar o *input* o programa também irá gerar um arquivo na extensão \*.gjf para que o usuário possa conferir se a geometria inicial do sistema a ser simulado está correta.

### B.3 - Lidando com o código Qd10 – GSA

A seguir estão descritos passo a passo as etapas de cálculos computacionais realizados pelo código Q<sub>d10</sub> - GSA:

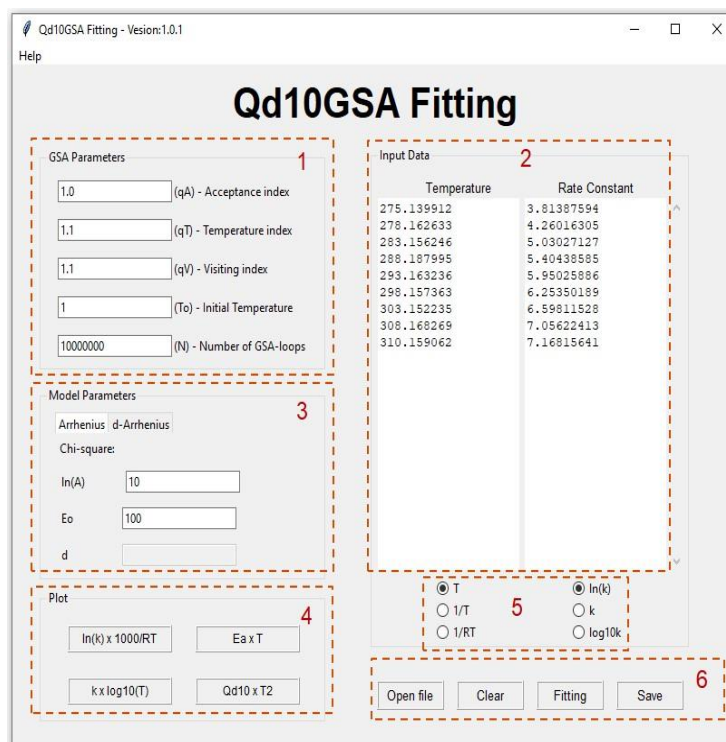
- Realizar o ajuste dos parâmetros das funções do modelo de *Arrhenius*, Equação (2), e do modelo proposto *d*-Arrhenius, Equação (12).
- Calcular o coeficiente usual da temperatura Q<sub>10</sub>, Equação (36).
- Calcular o coeficiente generalizado da temperatura Q<sub>d10</sub>, Equação (37);
- Calcular a energia de ativação, Equação **Erro! Fonte de referência não encontrada.**
- Realizar a construção de 4 opções de gráficos utilizando os resultados do ajuste

Na janela principal do código o usuário deve escolher entre duas opções: ajuste com o coeficiente usual Q<sub>10</sub> (modelo de *Arrhenius*) e o generalizado Q<sub>d10</sub> (modelo *d*-Arrhenius), ver Figura 30. Existem quatro subjanelas: (i) “*GSA Parameters*”, parâmetros usados no método de otimização estocástica; (ii) “*Model Parameters*”, local onde são exibidos os parâmetros das funções usadas nos modelos Q<sub>10</sub> e no Q<sub>d10</sub>; (iii) “*Plot*”, onde são disponibilizadas opções gráficas; e (iv) “*Input Data*”, local onde o usuário deve fornecer seus dados de entrada.

Os valores experimentais obtidos pelo usuário (dados de entrada para o programa) devem ser introduzidos nas duas colunas da sub janela “*Input Data*”, o usuário pode escolher entre digitar os valores diretamente nas colunas ou utilizar o botão “*Open File*” para selecionar um arquivo que contenha os dados (este arquivo deve estar em formato CSV ou ASCII-TXT, com opcional de cabeçalho). Além disso, o usuário deve informar o formato e as unidades de seus dados ao selecionar uma das opções *T*, *1/T* ou *1/RT* para a primeira coluna (referente a temperatura); e *ln(k)*, *k* ou *log<sub>10</sub> k* para a segunda coluna (referente a constante cinética). O usuário



deve clicar no botão “*Fitting*” para iniciar o processo de otimização e no botão “*Save*” para salvar os resultados ao fim do ajuste.



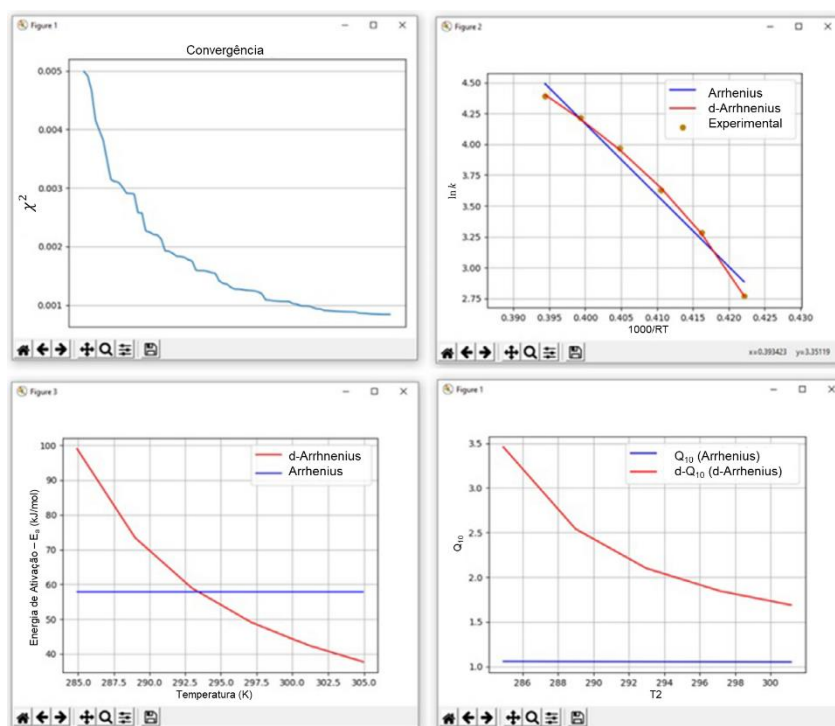
**Figura 30:** Janela principal do código Q<sub>d10</sub> - GSA. Esta tela mostra os dados de entrada do usuário bem como as ferramentas disponíveis no software. 1: Parâmetros do GSA; 2: Dados de entrada; 3: Parâmetros do modelo; 4: Ferramentas para a exibição de gráficos; 5: Formatos de entrada para temperatura e constante cinética; 6: Botões de ação.

Para utilizar corretamente a plataforma Q<sub>d10</sub> - GSA siga os seguintes passos:

- i. Para iniciar a interface execute o arquivo “*Qd10GSA\_FittingCode.pyw*” e a janela principal (Figura 30) irá abrir automaticamente com parâmetros sugeridos para o GSA e para os modelos de constante cinética.
- ii. Abra o arquivo contendo os dados experimentais ou digite os valores de temperatura e constante cinética da reação diretamente na tela de acordo com as opções de formatos e unidades de entrada. Se for selecionar um arquivo, certifique-se que ele esteja em uma das extensões: \*.dat, \*.txt ou \*.csv.
- iii. Selecione o modelo desejado (Arrhenius ou *d*-Arrhenius) para o ajuste dos dados experimentais. Nesta etapa, o código é inicializado automaticamente com parâmetros iniciais apropriados ao modelo. Uma sugestão dos autores é sempre realizar primeiramente o ajuste utilizando o modelo de Arrhenius, pois este modelo tem apenas 2 parâmetros e é mais facilmente ajustável. Deste modo, caso o usuário requeira o ajuste do

modelo *d*-Arrhenius posteriormente, os parâmetros encontrados para o modelo Arrhenius serão utilizados como parâmetros iniciais, facilitando e tornando mais rápido o processo de ajuste.

- iv. Para inicializar o processo de ajuste o usuário deve clicar no botão “*Fitting*”. O tempo do ajuste está diretamente relacionado ao número de iterações do algoritmo e pode ser alterado no campo “(N) – Number of GSA-loops”.
- v. Para melhorar o ajuste o usuário deve clicar novamente no botão “*Fitting*” ou aumentar número de *GSA-loops*.
- vi. Após o processo de otimização, é possível plotar diferentes tipos de gráficos clicando em algumas das opções da janela “*Plot*” (ver Figura 30 e Figura 31) ou salvar os parâmetros do ajuste em um arquivo de texto clicando no botão “*Save*”.



**Figura 31:** Exemplos dos gráficos que podem ser gerados através das opções contidas na sub janela “*Plot*”, após a realização do ajuste. No gráfico superior direito é possível visualizar os pontos que correspondem aos dados experimentais introduzidos no software, bem como as curvas dos modelos de Arrhenius e *d*-Arrhenius geradas através do ajuste.



**DECLARAÇÃO DE ORIGINALIDADE DE DISSERTAÇÃO DE MESTRADO OU TESE DE  
DOUTORADO**

Declaro que a presente dissertação/tese é original, elaborada especialmente para este fim, não tendo sido apresentada para obtenção de qualquer título e que identifico e cito devidamente todas as autoras e todos os autores que contribuíram para o trabalho, bem como as contribuições oriundas de outras publicações de minha autoria.

Declaro estar ciente de que a cópia ou o plágio podem gerar responsabilidade civil, criminal e disciplinar, consistindo em grave violação à ética acadêmica.

Brasília, 10 de Novembro de 2022

Assinatura do/a discente: Hugo Gontijo Machado

Programa: Programa de Pós Graduação em Química (PPGQ)

Nome completo: Hugo Gontijo Machado

Título do Trabalho: O uso da Inteligência Artificial e Outras Tecnologias  
na Caracterização de Fenômenos Químicos e Biológicos

Nível: ( ) Mestrado (x) Doutorado

Orientador/a: Kleber Carlos Mundim