



Universidade de Aveiro Departamento de Engenharia de Materiais e Cerâmica
2023

**Gabriel Luís Silva
Martins**

Aplicação de técnicas de aprendizagem automática para analisar a síntese de materiais lamelares nanoestruturados



**Universidade de
Aveiro
2023**

Departamento de Engenharia de Materiais e
Cerâmica

**Gabriel Luis Silva
Martins**

Aplicação de técnicas de aprendizagem automática para analisar a síntese de materiais lamelares nanoestruturados

Relatório de estágio apresentado à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia de Materiais, realizada sob a orientação científica do Doutor Tiago Galvão, Investigador Auxiliar, coorientação do Professor Doutor João Tedim, Professor Associado, ambos do Departamento de Engenharia de Materiais e Cerâmica da Universidade de Aveiro, e da Doutora Cláudia Rocha, da Smallmatek, Lda.

Este trabalho foi desenvolvido no âmbito do projeto FCT/MCTES DataCor (POCI-01-0145-FEDER-030256 e PTDC/QUI-QFI/30256/2017) e MSCE COAT4LIFE (GA 101007430).

O júri
Presidente

Prof. Doutora Paula Celeste da Silva Ferreira
Investigadora Coordenadora da Universidade de Aveiro

Doutor German Perez Sanchez
Investigador Doutoramento (nível 1) da Universidade de Aveiro

Doutor Tiago Luís Pereira Galvão
Investigador Auxiliar da Universidade de Aveiro

agradecimentos

Ao Doutor Tiago Galvão por toda a orientação, apoio, disponibilidade, paciência e por todo o conhecimento transmitido para a realização deste trabalho.

Ao Professor Doutor João Tedim pela disponibilidade, orientação e apoio para a elaboração deste trabalho.

À Doutora Cláudia Rocha e ao Doutor Frederico Maia, da Smallmatek, pelo fornecimento de dados e pela disponibilidade prestada.

À minha família, em especial, à minha mãe por tudo o que fez por mim, por todo o apoio incondicional e por me dar a oportunidade de completar a minha formação académica.

Aos meus amigos por todo o apoio emocional, pelo companheirismo e pelos momentos partilhados, ao longo destes anos, que me permitiram crescer, pessoalmente.

palavras-chave

Hidróxidos duplos lamelares, Intercalação, Inibidores de corrosão, Corantes, Fármacos, Difração de raios-X, Aprendizagem automática, Florestas aleatórias, Validação-cruzada, Descritores

resumo

Os hidróxidos duplos lamelares (HDL) pertencem a uma classe de materiais nanoestruturados sintéticos capazes de realizar troca iônica, que podem ser sintetizados de maneira relativamente barata, sendo estes materiais potencialmente recicláveis. Estes materiais são produzidos a nível industrial pela empresa Smallmatek, Lda. (SMT) e permitem o armazenamento e libertação controlada dos compostos intercalados, consoante a variação das condições do meio externo, permitindo dar diferentes funcionalidades quando incluídos como aditivos nanoestruturados em revestimentos.

Desta forma, a aplicação de técnicas de Aprendizagem Automática (AA) aos resultados experimentalmente obtidos permitirá à empresa minimizar o tempo e recursos financeiros necessários para o desenvolvimento de novos produtos, como, por exemplos, HDL capazes de intercalar inibidores de corrosão, biocidas/antimicrobianos, sensores, antocianinas, fármacos e/ou tensoativos.

Neste contexto, o principal objetivo deste trabalho foi prever se um determinado composto pode ser intercalado em HDL. Para tal, foi necessário obter e organizar o maior número de dados possível, considerando os dados obtidos, anteriormente na UA e na SMT.

Recorrendo a uma linguagem de programação de código aberto (Python), com este conjunto de dados foi possível obter um modelo de florestas aleatórias que considera diferentes tipos de condições experimentais. Após o tratamento de dados, utilizou-se o método de eliminação recursiva de descritores para remover iterativamente descritores menos importantes, criando um subconjunto de descritores que maximiza a métrica estatística selecionada. Deste modo, para validar a seleção de descritores efetuada pelo método de eliminação recursiva de descritores, foi realizada a validação cruzada. É importante avaliar a confiabilidade das próprias métricas de avaliação dos modelos de AA. Esta é um fator importante, pois pode ajudar a encontrar erros na avaliação e interpretação dos resultados. Assim, o valor de Kappa permite testar a confiabilidade das métricas estatísticas para avaliar o modelo. O coeficiente de correlação de Matthews, tal como o valor de Kappa, permite avaliar a confiabilidade do modelo, sendo a sua interpretação semelhante ao valor de Kappa.

Através dos resultados obtidos recorrendo à validação cruzada usando os descritores selecionados, verifica-se que a exatidão, exatidão balanceada, sensibilidade e precisão, estão todos acima de 70 %, o que já permite prever corretamente a maioria das sínteses. Para o teste final foi possível verificar que os resultados apontam para uma melhoria de todos os parâmetros, exceto para a sensibilidade, provando-se, desta forma, que o modelo teve um bom desempenho para novos dados, e que nunca foram usados na escolha dos descritores e otimização dos seus parâmetros.

keywords

Layered double hydroxides, Intercalation, Corrosion inhibitors, Dyes, Drugs, X-ray diffraction, Machine Learning, Random Forests, Cross-validation, Descriptors.

abstract

Layered double hydroxides (LDHs) belong to a class of synthetic nanostructured materials capable of performing ion exchange, which can be synthesized in a relatively cheap and potentially recyclable way. These materials are produced at an industrial level by the company SMT and can store and release in a controlled way the intercalated compounds, depending on changes in external environmental conditions, providing different functionalities when included as nanostructured additives in coatings, for example. In this way, applying Machine Learning techniques to the results experimentally obtained will allow the company to minimize the time and financial resources needed to develop new products, such as corrosion inhibitors, biocides/antimicrobials, sensors, anthocyanins, drugs and/or surfactants. In this context, the main objective of the present work was to predict whether a given compound can be intercalated in LDHs. To achieve this, it was necessary to obtain and organize as much data as possible, considering data obtained experimentally, both in this work and in previous ones, by UA and SMT. In this way, a large amount of data allowed us to obtain a more assertive and reliable model.

With this data set, it was possible to obtain a random forest model that considers different types of experimental conditions, using an open-source programming language (Python), which could enable the construction of applications in future work. After data processing, the recursive descriptor elimination method was used to iteratively remove less important descriptors, creating a subset of descriptors that maximizes the selected statistical metric. To validate the selection of descriptors carried out by recursive feature elimination method, cross-validation was performed. It is important to assess the reliability of AA models' own evaluation metrics. This is an important factor as it can help to find errors in the evaluation and interpretation of results. Thus, the Kappa value allows testing the reliability of statistical metrics to evaluate the model. The Matthews correlation coefficient, like the Kappa value, allows the reliability of the model to be assessed, with its interpretation being similar to the Kappa value.

Through the values obtained with cross-validation with the selected descriptors, it can be seen that the accuracy, balanced accuracy, sensitivity and precision are all above 70%, which already allows the majority of synthesis to be correctly predicted. For the final test, it was possible to verify that the results point to an improvement in all parameters, except sensitivity, thus proving that the model performed well for new data, and that they were never used in the choice of descriptors and optimization of their parameters.

Índice

Capítulo 1 - Introdução.....	1
1.1 - Enquadramento do trabalho.....	1
1.2 - Hidróxidos duplos lamelares.....	2
1.3 - Aplicações de HDL.....	4
1.4 - Síntese.....	9
1.5 - Estrutura, Morfologia, e outras propriedades.....	11
1.7 - Aprendizagem automática em Ciência e Engenharia dos Materiais.....	13
1.8 - Inovações do trabalho em relação ao estado da arte.....	16
Capítulo 2 - Detalhes experimentais.....	18
2.1 - Materiais usados.....	18
2.2 - Síntese do material precursor (HDL).....	20
2.3 - Intercalação dos compostos em HDL.....	20
2.4 - Lavagem e secagem.....	22
2.5 - Análise por difração raios-X.....	23
Capítulo 3 - Aprendizagem automática.....	24
3.1 - Passos para a criação e utilização do modelo AA.....	24
3.1.1 - Base de dados.....	24
3.1.2 - Tratamento de dados.....	26
3.1.3 - Seleção dos descritores.....	26
3.1.4 - Métodos de AA: Otimização do modelo.....	29
3.1.5 - Validação e teste do modelo, e métricas estatísticas.....	31
3.1.6 - Importância dos descritores pelo método de permuta.....	35
Capítulo 4 - Apresentação e discussão de resultados.....	36
4.1 - Resultados gerais das intercalações.....	36
4.2 - Análise dos resultados de intercalação parcial.....	42
4.3 - Análise dos resultados de aprendizagem automática.....	48
4.4 - Importância dos descritores.....	50
Capítulo 5 - Conclusões e trabalho futuro.....	52
Referências Bibliográficas.....	55

Índice de Figuras

Figura 1 - Esquema da criação da base de dados até à aplicação de um modelo de AA para prever se um determinado composto funcional pode ser intercalado em materiais nanoestruturados baseados em hidróxidos duplos lamelares (HDL).	2
Figura 2 - Representação esquemática da estrutura de um hidróxido duplo lamelar tradicional genérico (HDL) representada lateralmente e ao longo do eixo cristalográfico <i>c</i> , representando as folhas octaédricas distorcidas. Adaptado de [3].	3
Figura 3 - Mecanismo de troca de inibidores com iões cloreto para proteção da superfície[26].	9
Figura 4 - Representação da síntese de HDL por co-precipitação[27].	10
Figura 5 - Esquema troca iónica[28].	10
Figura 6 - Representação do processo de síntese envolvendo co-precipitação para obtenção das primeiras partículas de HDL, seguido de tratamento hidrotérmico para promover a sua cristalinidade	11
Figura 7 - Padrão Zn(2)Al-NO ₃ XRD e reflexões mais relevantes da amostra (pH = 10, T = 373 K, t = 4 h). Retirado de [1].	12
Figura 8 - Exemplo de uma placa hexagonal de HDL obtida por MFA para uma amostra sintetizada por co-precipitação nas seguintes condições: pH = 10, T = 373 K, t = 4 h. Retirado de [1].	13
Figura 9 - Equipamento de centrifugação	23
Figura 10 - Imagem parcial da base de dados.	25
Figura 11 - Representação do modelo de eliminação recursiva de descritores (Adaptado [41]).	29
Figura 12 - Representação da execução do método de árvores de decisão.	30
Figura 13 - Representação da execução do método de agregação. Retirado de [43].	31
Figura 14 - Representação da execução do método das florestas aleatórias. Retirado de	31
Figura 15 - Esquema de validação cruzada.	32
Figura 16 - Matriz de confusão, onde a classificação do sucesso da intercalação resulta em 4 possibilidades diferentes.	33
Figura 17 - DRX do HDL-NO ₃ e do HDL-citrato; b) Molécula de citrato.....	40
Figura 18 - a) DRX do HDL-NO ₃ e do HDL-BTA (pH = 8,1); b) Molécula de BTA desprotonada.	41
Figura 19 - a) DRX do HDL-NO ₃ e do HDL-BTA (pH = 10,0); b) Molécula de BTA desprotonada.	42
Figura 20 - a) Molécula de ácido ascórbico; b) Anião da molécula de ácido ascórbico; c) Dianião da molécula de ácido ascórbico.	45
Figura 21 - a) DRX do HDL-NO ₃ e HDL-ascorbato (pH = 8,0).	43
Figura 22 - a) Difratoograma de raios-X do HDL-NO ₃ e HDL-ascorbato (pH = 8,7).	44
Figura 23 - a) Difratoograma de raios-X do HDL-NO ₃ e HDL-ascorbato (pH = 10,0); b) Molécula ácido ascórbico desprotonada.....	44
Figura 24 - Correlação entre pH e espaçamento basal do ácido ascórbico.....	45
Figura 25 - a) Difratoograma de raios-X do HDL-NO ₃ e HDL-ácido salicílico (pH = 8,2); b) Molécula ácido salicílico desprotonado.	46
Figura 26 - a) Difratoograma de raios-X do HDL-NO ₃ e HDL-ácido salicílico (pH = 9,4); b) Molécula ácido salicílico desprotonado.	46

Figura 27 - a) Difratoograma de raios-X do HDL-NO ₃ e HDL-sulfamato (pH = 8); b) Molécula ácido sulfâmico desprotonado.....	47
Figura 28 - a) Difratoograma de raios-X do HDL-NO ₃ e HDL-sulfamato (pH = 10,2); b) Molécula ácido sulfâmico desprotonado.....	48
Figura 29 - Diagrama de importância de descritores avaliada pelo método da permuta de descritores. ...	51

Índice de tabelas

Tabela 1 - Compostos usados na intercalação em HDL.....	18
Tabela 2 - Valores de pKa e pH dos compostos.....	21
Tabela 3 - Interpretação dos valores de Kappa.....	35
Tabela 4 - Principais condições de intercalação, resultados das diferentes intercalações, e distância entre camadas.....	36
Tabela 5 - Resultados obtidos na validação e teste do modelo.....	49

Lista de Abreviaturas

AA - Aprendizagem automática

HDL - Hidróxido duplo lamelar

DEMaC - Departamento de Engenharia dos Materiais e Cerâmica

SMT - Smallmatek, Lda

DRX - Difração de raios-X

VO_x - Vanadato

SECOP - *Surface Engineering and Corrosion Protection*

MFA - Microscopia de força atômica

RQEP - Relação Quantitativa entre Estrutura e Propriedade

MVR - Máquina vetorial de relevância

ASE - Área superficial específica

RSV - Regressão de suporte vetorial

d - Espaçamento basal

ERD - Eliminação recursiva de descritores

CMM - Coeficiente de correlação de Matthews

VP - Verdadeiros positivos

VN - verdadeiros negativos

FP - Falsos positivos

FN - Falsos negativos

BTA - Benzotriazo

Capítulo 1 - Introdução

1.1 - Enquadramento do trabalho

A aprendizagem automática (AA) é um ramo da inteligência artificial que permite que um determinado programa seja bastante exato na previsão de resultados, mesmo sem ser expressamente programado para isso.

As aplicações baseadas em AA recorrem a algoritmos que conseguem prever propriedades recorrendo a um conjunto de dados, enquanto uma aplicação expressamente programada para calcular uma determinada propriedade recorre normalmente a equações já conhecidas para esse cálculo.

A inteligência artificial possibilita às máquinas aprender com as experiências, enquanto vão ajustando as previsões com novas entradas de dados e por fim, idealmente, desempenha tarefas como os humanos. Esta ferramenta possibilita longas horas de trabalho contínuo e, desta forma, permite inovar em vários setores de trabalho.

Neste trabalho híbrido experimental/computacional, pretende-se recorrer à AA para produzir um modelo que permita prever se um determinado composto funcional pode ser intercalado em materiais nanoestruturados baseados em hidróxidos duplos lamelares.

Estes materiais são produzidos a nível industrial pela empresa SMT e permitem o armazenamento e libertação controlada dos compostos intercalados consoante variação das condições do meio externo, permitindo dar diferentes funcionalidades quando, por exemplo, incluídos como aditivos nanoestruturados em revestimentos.

Assim, a aplicação de técnicas de AA aos resultados obtidos permitirá à empresa minimizar o tempo e recursos financeiros necessários para o desenvolvimento de novos produtos, como HDL intercalados com inibidores de corrosão, biocidas/antimicrobianos, sensores, antocianinas, fármacos e/ou tensoativos.

No esquema da Figura 1 está representado o progresso a ser feito neste estágio para se conseguir criar a base de dados a partir da preparação de HDL por troca iónica, produzindo assim um modelo de AA para se conseguir prever os resultados.

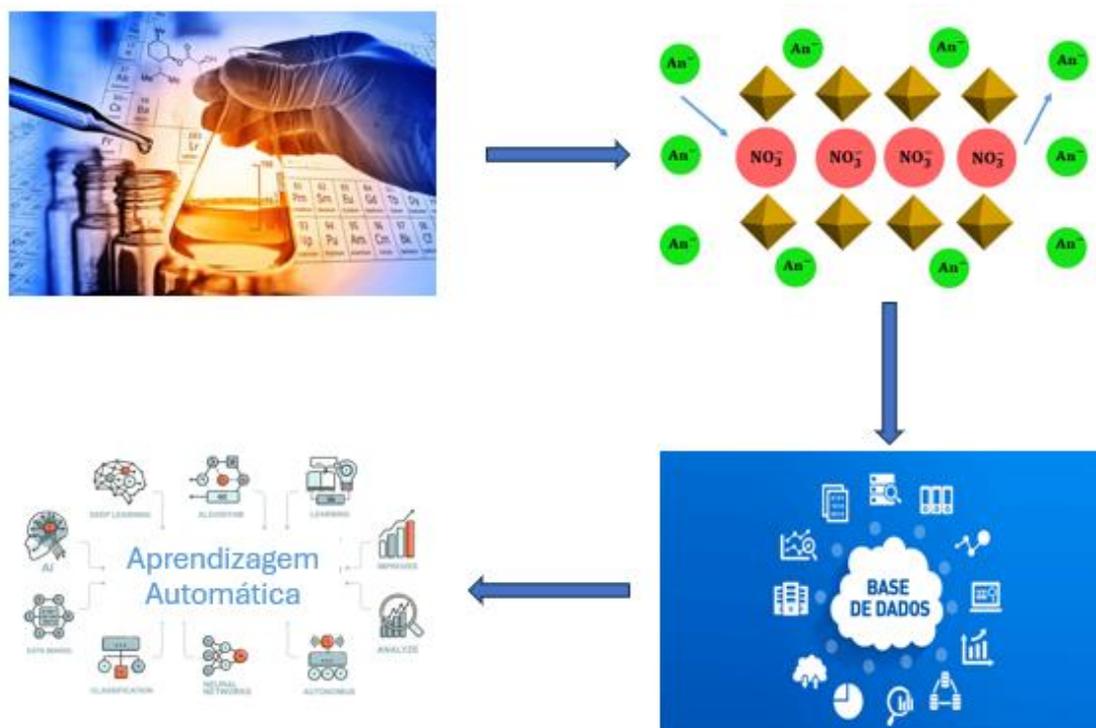


Figura 1 - Esquema da criação da base de dados até à aplicação de um modelo de AA para prever se um determinado composto funcional pode ser intercalado em materiais nanoestruturados baseados em hidróxidos duplos lamelares.

1.2 - Hidróxidos duplos lamelares

Hidróxidos duplos lamelares pertencem a uma classe de materiais nanoestruturados capazes de realizar troca iónica, e que podem ser sintetizados de maneira relativamente barata e potencialmente reciclável.

As composições mais comuns de HDL reportadas na literatura científica apresentam a fórmula geral $[M^{2+}_{1-x}M^{3+}_x(OH)_2]^{x+}(A^{n-})_{x/n} \cdot mH_2O$, onde M^{2+} e M^{3+} são catiões di- e trivalentes, respetivamente, A^{n-} são aniões e x é igual à razão molar $M^{3+} / (M^{2+} + M^{3+})$ que normalmente está compreendida entre 0,22 e 0,33 [1,2].

Através da combinação de diferentes catiões e aniões usando diversos métodos de síntese é possível obter um elevado número de composições de HDL com propriedades físicas e químicas desejáveis para cada aplicação [1,2]. Um exemplo para a formação e observação da estrutura da camada catiónica de HDL, é argila do tipo brucite (brucite = $Mg(OH)_2$), onde esta é constituída por camadas catiónicas coordenadas octaetricamente com OH^- , podendo estes serem ligeiramente distorcidos devido aos octaedros de metais diferentes que os rodeiam. A carga positiva das camadas catiónicas

resulta da substituição parcial de Mg por Al (no caso de HDL baseados em Mg_2Al , por exemplo), sendo estas cargas equilibradas pelos aniões das intercamadas e moléculas de água, permitindo que as camadas se empilhem umas sobre as outras através de forças maioritariamente eletrostáticas e pontes de hidrogénio [3].

A Figura 2 mostra um esquema de uma estrutura genérica de um HDL com razão molar $M^{2+}:M^{3+}$ de 2:1, sendo que em muitos minerais existentes na natureza (por exemplo, a hidrotalcite) estes são intercalados com aniões carbonato hidratados. As posições que os octaedros M^{3+} e M^{2+} adotam na camada são bem definidas na rede cristalina, embora dependam da razão entre as espécies trivalentes e divalentes [3].

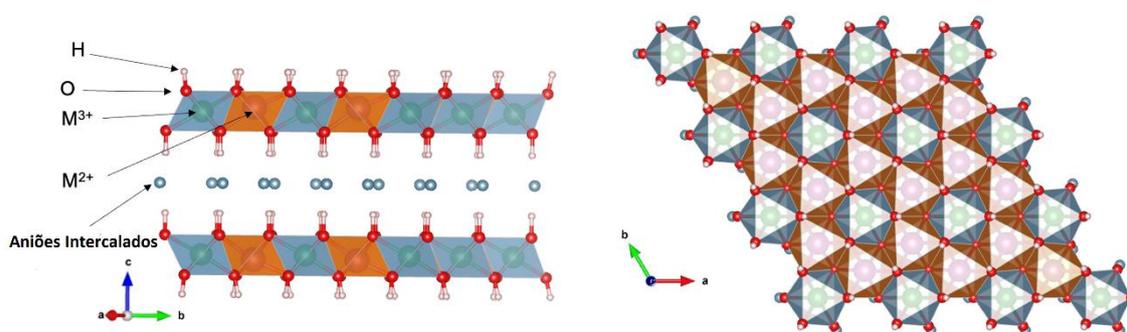


Figura 2 - Representação esquemática da estrutura de um hidróxido duplo lamelar tradicional genérico representada lateralmente e ao longo do eixo cristalográfico c , representando as folhas octaédricas distorcidas. Adaptado de [3].

Os HDL têm uma estrutura tipo sanduíche na qual os aniões inorgânicos ou orgânicos estão intercalados entre as camadas de metal carregadas positivamente de maneira repetida. Uma propriedade particularmente importante dos HDL é a sua capacidade de trocar aniões nas intercamadas, sendo assim conhecidas também como argilas aniónicas. As camadas de hidróxidos metálicos podem ser fabricadas com a combinação de diferentes iões metálicos divalentes (Zn^{2+} , Mg^{2+} , Mn^{2+} , Fe^{2+} , Pb^{2+}) e trivalentes (Al^{3+} , Cr^{3+} , Fe^{3+}). As características estruturais dos HDL permitem a possibilidade de usar uma grande variedade de compostos intercalados, seja por modificação da composição química da camada de hidróxido ou por modificação química/estrutural das intercamadas. O tamanho da partícula lamelar obtida está compreendido entre algumas dezenas de nanómetros (nm), podendo ir até a alguns micrómetros (μm), dependendo das condições de síntese [2,3].

1.3 - Aplicações de HDL

Os HDL têm uma estrutura única e propriedades que podem ser manipuladas para diferentes fins, desde o desenvolvimento de catalisadores, supercondensadores, nanocompósitos polímero/HDL com propriedades mecânicas melhoradas, materiais para troca aniônica, absorventes de CO₂ entre outras substâncias (*i.e.* poluentes) e retardantes de chama [2,4].

As ligações relativamente fracas entre as camadas, combinada com a natureza lábil dos aniões intercalados, confere aos HDL versatilidade na forma como pode alocar diferentes espécies químicas nas intercamadas, abrindo assim inúmeras possibilidades à sua aplicação como materiais para troca iônica com uma multiplicidade de composições químicas. Tem havido um interesse considerável no uso de HDL para remover espécies carregadas negativamente através de adsorção na superfície e/ou de troca aniônica.

A grande capacidade dos HDL de captar espécies aniônicas pode ser explicada pela sua elevada área de superfície, assim como a sua alta capacidade de efetuar troca iônica [5]. Esta capacidade de efetuar troca iônica deve-se ao facto do espaço interlamelar ser flexível, sendo acessível a espécies moleculares polares, mas sobretudo a espécies carregadas negativamente. A capacidade de troca aniônica dos HDL é afetada pela natureza dos aniões interlamelares inicialmente presentes e pela densidade de carga da camada (ou seja, a razão molar M(II)/M(III) nas camadas catiónicas). Em particular, a entalpia favorável de estabilização da estrutura dos HDL associada ao CO₃²⁻ resulta nesses aniões serem difíceis de retirar do interior das camadas através de reações de troca aniônica.

Os HDL podem fixar espécies aniônicas da solução através de três mecanismos: adsorção na superfície, troca aniônica interlamelar e reconstrução de um HDL calcinado recorrendo a um chamado “efeito memória”. O “efeito memória” dos HDL é uma das suas características mais singulares para serem usados como adsorventes de espécies aniônicas. A calcinação permite a reciclagem e reutilização dos materiais após realizada a adsorção de espécies, uma vez que permite a eliminação dos aniões captados sob a forma de dióxido de carbono e água, podendo a estrutura do HDL ser reconstruída após calcinação através da hidratação em água [6,7].

O caso da aplicação de HDL para catálise, deve-se sobretudo à possibilidade de obter óxidos metálicos mistos a partir de HDL por decomposição térmica controlada, resultando em materiais com áreas de superfície específicas elevadas (100-300 m² g⁻¹),

acompanhados de reatividade da superfície propícia para a catálise, numa dispersão homogênea e termicamente estável dos vários componentes químicos baseados nos iões metálicos mistos, e na possibilidade de reconstrução da estrutura em condições de pH próximo do neutro [5].

As primeiras aplicações médicas dos HDL foram principalmente como agentes antiácidos e anti-tripepsina e espera-se um aumento da procura nessa área no futuro. Além disso, também têm sido sugeridos para a remoção de aniões de fosfato do fluido gastrointestinal com o objetivo de prevenir a hiperfosfatemia, uma doença relacionada com níveis baixos de fósforo no sangue [5]. Mais recentemente, os HDL encontraram outras aplicações significativas na medicina, especialmente em formulações farmacêuticas. Estudos recentes têm-se focado na intercalação e libertação controlada de compostos farmacologicamente ativos a partir de HDL, aproveitando a sua biocompatibilidade, composição química variável, capacidade de intercalar fármacos aniônicos e o seu caráter alcalino.

As tendências atuais na tecnologia farmacêutica exigem que as formulações sejam capazes de manter os níveis de fármaco farmacologicamente ativos por longos períodos, evitando a administração repetida e/ou libertar localmente o fármaco num determinado alvo farmacêutico [5]. A região interlamelar dos HDL pode ser considerada um nanoreservatório no qual o fármaco é armazenado de forma ordenada, mantendo a sua integridade e protegendo-o da ação da luz e do oxigénio. Após a administração do material com o composto intercalado, o fármaco pode ser libertado por troca iónica. Este sistema de libertação de fármacos pode reduzir os seus efeitos colaterais associados a concentrações elevadas no sangue, prolongando, ao mesmo tempo, a vida útil do princípio ativo. A taxa de difusão do fármaco para fora do HDL é controlada pela força da interação hospedeiro-hóspede, pelos gradientes de concentração das espécies aniónicas no exterior, a rigidez das camadas e o comprimento do caminho de difusão definido pela extensão das camadas [5].

No que toca a aplicações eletroquímicas, a capacidade dos HDL trocarem iões permite obter elétrodos modificados com este tipo de argilas que têm sido extensivamente estudados, tais como biossensores [8].

Quanto aos nanocompósitos de polímero/HDL, a maior aplicação atual de materiais baseados em HDL é na indústria de polímeros, principalmente para estabilizar o policloreto de vinilo [9,10]. Isto permite substituir outras metodologias baseadas em metais pesados que são perigosos para o meio ambiente.

Aplicações de HDL em corrosão

No DEMaC e na SMT encontramos exemplos de trabalhos de uma das principais aplicações de HDL, que é na proteção contra a corrosão. Na tentativa de conferir proteção anticorrosiva a substratos metálicos, é comum usar-se revestimentos orgânicos protetores com compostos iônicos e moleculares (inibidores de corrosão). O revestimento funciona como uma barreira contra a entrada de água, oxigênio e outras espécies (proteção passiva). No entanto, com a degradação do revestimento quando exposto a condições de serviço (ex.: radiação eletromagnética, gradientes de temperatura, ...) o efeito barreira diminui, sendo necessário que os inibidores de corrosão atuem diretamente na superfície do metal, reduzindo a velocidade corrosão (proteção ativa). Entre os compostos mais utilizados e bem-sucedidos neste efeito de proteção ativa encontram-se os compostos derivados do Cr(VI). No entanto, por razões de elevada toxicidade, a utilização de cromatos foi proibida, havendo uma tentativa de substituição de Cr(VI) por aditivos “inteligentes” que consigam demonstrar um desempenho semelhante ou até superior aos compostos de Cr(VI). Estes novos aditivos consistem em materiais que contêm inibidores de corrosão, e que libertam os inibidores de corrosão de forma controlada, limitando a sua lixiviação e interação negativa com o revestimento [3]. Entre os mais promissores estão os HDL.

Uma das razões para se escolher HDL para a proteção contra a corrosão, em detrimento de outros nanocontenores está relacionado com o fato de os HDL poderem ser usados tanto na forma de pigmentos anticorrosivos incorporados num revestimento, bem como parte de um filme de conversão diretamente no substrato metálico a proteger [11–13]. Isso significa que ele pode ser facilmente adaptado de acordo com diferentes parâmetros. Consequentemente, os HDL podem ser usados como um pó/pasta para conferir propriedades de proteção ativa adicionais a um revestimento de barreira pela adição de uma funcionalidade de autorreparação [14].

A vasta gama de trabalhos relativos a materiais à base de HDL, como partículas (nanocontenores) ou revestimentos de conversão, começou com trabalhos pioneiros de Buchheit [15,16], Williams e McMurray [17,18], em meados dos anos de 1990 e início dos anos 2000. Tanto os revestimentos de conversão como os pigmentos ativos de proteção anticorrosiva à base de HDL visam substituir o Cr hexavalente, principalmente, em aplicações onde os pré-tratamentos e primários à base de Cr(VI) foram a principal solução para proteção anticorrosiva na indústria aeroespacial [19]. Assim, não é

surpreendente que estes primeiros trabalhos se tenham concentrado maioritariamente em ligas de alumínio.

Em relação ao uso de HDL como pigmentos anticorrosivos em revestimentos orgânicos, Williams e McMurray investigaram o HDL-Mg-Al incluindo nitrato, carbonato e cromato, que foram adicionados a um revestimento de polivinilbutiral aplicado sobre substratos de liga de alumínio 2024-T3 para investigar como esses pigmentos podem afetar o desenvolvimento da corrosão [17]. Em todos os revestimentos modificados com HDL, as taxas de degradação do revestimento foram reduzidas, tendo-se verificado que a eficiência inibidora dos HDL depende da natureza do anião intercalado, com o HDL contendo cromato exibindo o melhor desempenho.

Num estudo contemporâneo independente de Buchheit e seus colegas, a preparação de HDL-Zn-Al intercalado com decavanadato foi descrita pela primeira vez e o efeito desse material foi investigado para um revestimento de resina epóxi. O HDL-Zn-Al resultante foi preparado primeiro modificando a especiação do vanadato (VO_x) da forma meta^- para decavanadato e, em seguida, obteve-se o HDL por co-precipitação direta numa solução descarbonatada. A análise estrutural realizada por difração de raios-X (DRX) mostrou a expansão das intercamadas quando o decavanadato estava presente nas intercamadas dos HDL, em relação ao HDL contendo cloreto. Estudos de libertação realizados em solução de NaCl 0,5 M, revelaram a libertação de espécies de Zn e VO_x , sendo esta última atribuída à reação de troca aniônica entre cloretos e vanadatos. Por outro lado, a libertação de Zn foi associada à co-intercalação de Zn^{2+} ou à formação de um complexo metálico entre Zn e VO_x , que posteriormente foi imobilizado nas intercamadas do HDL. Adicionalmente, ensaios de corrosão demonstraram um claro efeito protetor dos revestimentos carregados com HDL contendo vanadato (testes de nevoeiro salino e espectroscopia de impedância eletroquímica). Aliás, o mesmo trabalho descreveu a possibilidade de usar HDL como pigmento sensor de corrosão, porque a análise DRX de HDL-vanadato (HDL-VO_x) exposta à solução de NaCl, diretamente como pó ou embutido na resina epóxi, revela a ocorrência de uma segunda fase de HDL associada à intercalação de cloretos, constituindo uma forma de detetar a absorção mais precoce do eletrólito pelo revestimento (sensor de corrosão) [15].

Como mencionado anteriormente, os trabalhos do grupo de Buchheit não se restringiram ao uso de HDL como pigmento em revestimentos orgânicos. Eles também lançaram as bases para alguns dos trabalhos mais populares em termos de proteção ativa, sem Cr, pré-tratamentos de superfície e pigmentos à base de Li para revestimentos

orgânicos [20]. Nos seus trabalhos, Buchheit e os colegas usaram diferentes condições químicas de banho oxidante para obter HDL-Li-Al [16], na tentativa de reduzir os tempos de processamento para formação de revestimento e aumentar a resistência à corrosão. Os resultados mostraram que as composições químicas dos HDL baseados em nitrato/persulfato deram origem a uma melhor resistência à corrosão quando comparadas às composições baseadas, somente, em carbonato ou nitrato.

Após estes estudos iniciais, vários outros têm vindo a ser realizados combinando diversas formulações de HDL e diferentes inibidores de corrosão, onde se tem destacado a nível internacional os estudos realizados pelo grupo de investigação SECOP (*Surface Engineering and Corrosion Protection*) do DEMaC, conforme foi evidenciado num artigo de revisão publicado recentemente [3].

Em relação ao mecanismo de atuação dos HDL em si, os inibidores de corrosão na forma aniónica podem ser intercalados em HDL para aplicação na proteção contra a corrosão e libertados a partir destes materiais em condições ácidas ou alcalinas [21]. A pH elevado, os inibidores são trocados por aniões de hidróxido nas intercamadas, enquanto a pH mais baixo as camadas catiónicas dos HDL se dissolvem, libertando os inibidores nas intercamadas. Este mecanismo torna estes materiais tampões de pH, contribuindo assim, também, entre outros, para minimizar os efeitos da corrosão relacionados com a variação local de pH. Ademais, a troca aniónica não se restringe aos aniões hidróxido a partir das mudanças de pH. Os HDL também são capazes de libertar inibidores enquanto captam aniões agressivos, como, por exemplo, cloretos [22], tornando os HDL mais versáteis do que a maioria dos nanocontentores para aplicações anticorrosivas, como demonstra a Figura 3[23].

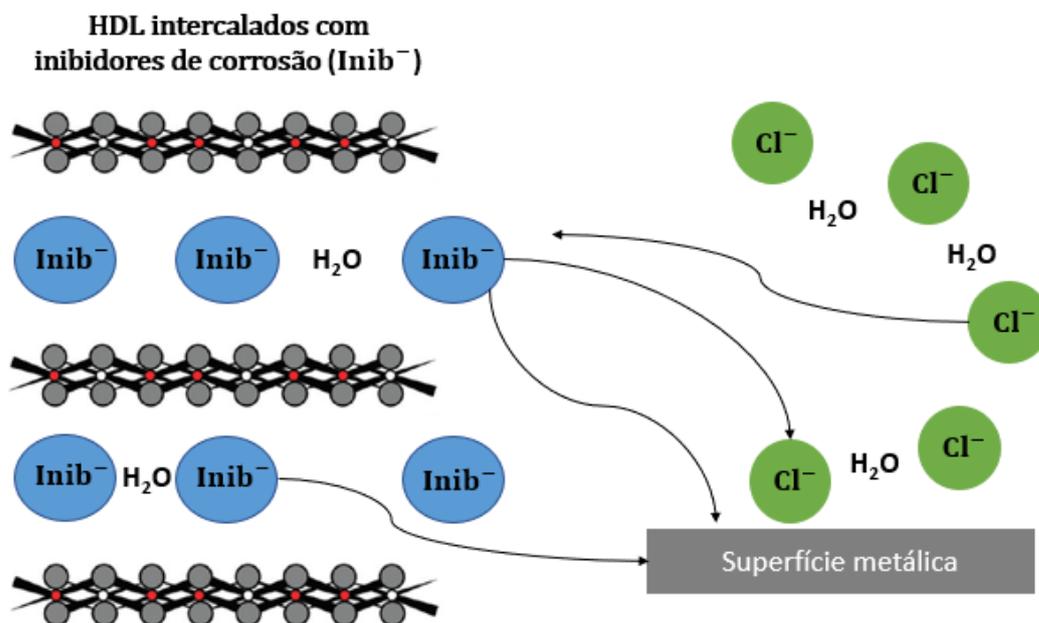


Figura 3 - Mecanismo de troca de inibidores com íons cloreto para proteção da superfície [24].

1.4 - Síntese

Dependendo da aplicação, existem vários métodos que podem ser usados para obter pós de HDL, baseando-se sobretudo em reações químicas, tais como, co-precipitação, métodos sol-gel, método de hidrólise de uréia, método hidrotérmico, reformação ou até moagem mecânica.

O método de co-precipitação, ilustrado na Figura 4 é o mais direto e comumente aplicado, sendo que será o método usado para obter o HDL usado como precursor neste trabalho (HDL-NO_3). Este método pode envolver tão somente a síntese numa só etapa, ou ser seguido por reações de troca aniônica, entre o anião intercalado e o anião exterior, possibilitando a intercalação de aniões orgânicos funcionais, entre outros, e que será implementada experimentalmente na SMT para se obterem novos materiais, como ilustrado na Figura 5 [3].

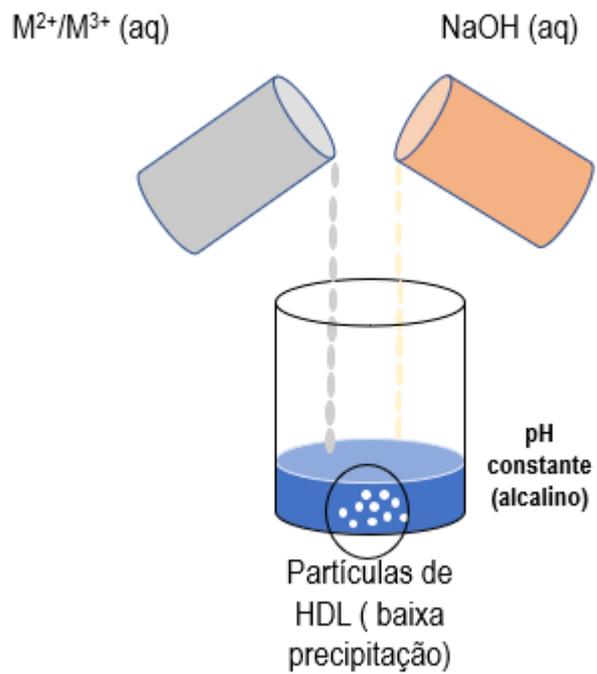


Figura 4 - Representação da síntese de HDL por co-precipitação [25].

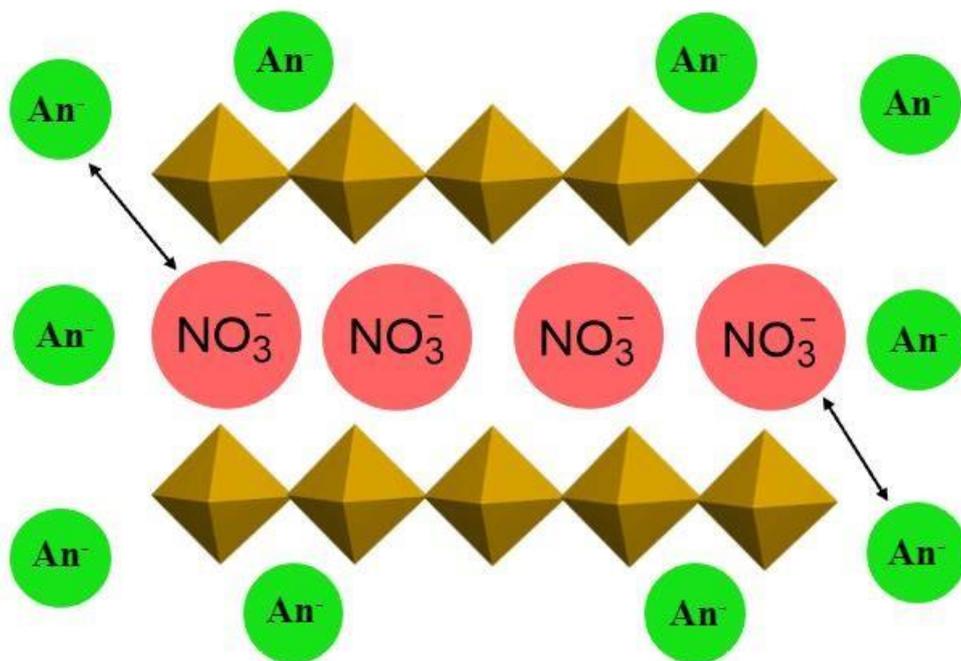


Figura 5 - Esquema troca iónica [26].

A síntese por co-precipitação pode ser realizada em condições de pH variável (co-precipitação por titulação) ou a pH constante. A última opção é preferível para obter HDL puros e cristalinos. Durante a reação, o pH da solução é mantido constante pela adição simultânea de uma solução alcalina (por exemplo, NaOH ou NH₄OH) juntamente com a solução precursora de sais metálicos mistos (metais que farão parte do HDL). Normalmente, uma solução alcalina é escolhida de acordo com os sais metálicos correspondentes e o anião desejado para ser intercalado nas intercamadas de HDL. Além disso, como é difícil evitar a presença de CO₂ no ar, é ainda aconselhável descarbonatar previamente todas as soluções (levando à ebulição e saturando com N₂ ou Ar) e seguidamente trabalhar sob fluxo dos gases inertes acima referidos para evitar a formação de estruturas HDL com carbonato intercalado, se estas não forem desejáveis [3].

O produto de HDL obtido na síntese por co-precipitação depende do controlo do pH do meio reacional (normalmente, alcalino), da concentração e natureza das soluções alcalinas e precursoras de metais (além da razão molar dos cátions metálicos), da temperatura e do tempo de tratamento hidrotérmico que vai promover o crescimento das partículas de HDL, conforme está representado na Figura 6 [11].

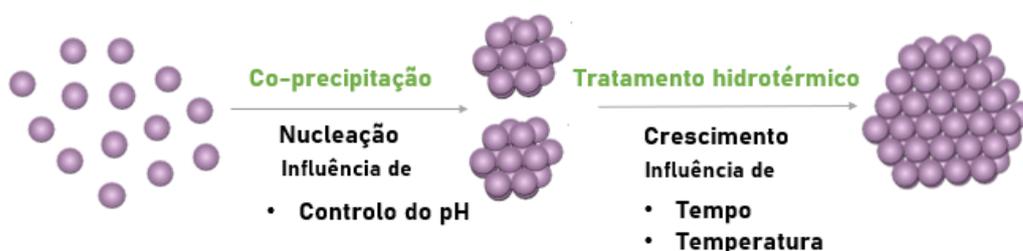


Figura 6 - Representação do processo de síntese envolvendo co-precipitação para obtenção das primeiras partículas de HDL, seguido de tratamento hidrotérmico para promover a sua cristalinidade

1.5 - Estrutura, Morfologia, e outras propriedades

Nesta secção apresenta-se algumas das principais características de HDL, no exemplo específico do HDL de zinco-alumínio, uma vez que é uma das composições mais utilizadas para intercalação de diferentes aniões e será o principal material usado no presente trabalho.

Os HDL de zinco-alumínio com nitrato intercalado ($\text{Zn}(n)\text{Al-NO}_3$, $n = \text{Zn}/\text{Al}$) são um material intermediário para a intercalação de diferentes moléculas funcionais usadas em várias aplicações industriais.

Entre os vários trabalhos já descritos na literatura, destaca-se aqui um trabalho onde a síntese de $\text{Zn}(2)\text{Al-NO}_3$ foi investigada pelo grupo do DEMaC, onde se varia as condições de síntese (pH, condições de tratamento hidrotérmico) [1]. Os padrões de DRX das amostras obtidas são típicos de um $\text{Zn}(2)\text{Al-NO}_3$ monofásico, conforme se verificou noutros estudos já reportados na literatura [21,24]. Como exemplo, a Figura 7 mostra o padrão de DRX de uma amostra sintetizada com as seguintes condições: pH = 10, $T = 373 \text{ K}$, $t = 4 \text{ h}$. Os picos mais intensos com ângulos 2θ iguais a $10,0$, $20,1$ e $30,0^\circ$ podem ser associados aos planos de reflexão (003), (006) e (009) (a atribuição dos picos restantes a uma fase de $\text{Zn}(2)\text{Al-NO}_3$, foi realizada em trabalhos anteriores [27].

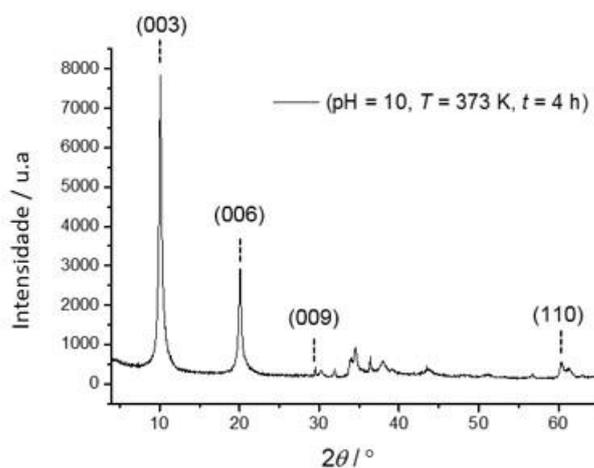


Figura 7 - Padrão de DRX de $\text{Zn}(2)\text{Al-NO}_3$ e reflexões mais relevantes da amostra (pH = 10, $T = 373 \text{ K}$, $t = 4 \text{ h}$). Retirado de [1].

A partir da posição da reflexão (003), Galvão *et al.*, calcularam um espaçamento basal d de $0,877 \text{ nm}$, que inclui a espessura de uma camada catiónica e uma intercamada. A partir da relação $c = 3d(003)$, estes autores conseguiram calcular o parâmetro de rede $c = 2,631 \text{ nm}$, o que está de acordo com os valores da literatura ($c = 2,675 \text{ nm}$ [35] ou $c = 2,667 \text{ nm}$ [28]).

Um pico correspondente à reflexão (110) também foi detetado. Esta reflexão é independente do empilhamento da camada e é determinada pelo tamanho e proporção

dos cátions metálicos nas camadas catiónicas. Da posição desta reflexão foi possível obter um espaçamento d correspondente a 0,1532 nm, que permite calcular o parâmetro de rede cristalina $a = 0,3064$ nm, segundo $a = 2d$ (110) [1] também semelhante a outros estudos de Zn-Al-NO₃ com Zn:Al ≈ 2 ($a = 0,3067$ nm [27] e 0,3077 nm [28]). A proporção metálica na amostra ($pH = 10$, $T = 373$ K, $t = 4$ h) foi obtida por espectroscopia de emissão atômica com plasma acoplado indutivamente como sendo $(1,8 \pm 0,2)$ Zn para $(1,0 \pm 0,2)$ Al.

Posteriormente, com ajuda de microscopia de força atômica (MFA) caracterizaram amostra ($pH = 10$, $T = 373$ K, $t = 4$ h), indicando a morfologia típica da placa hexagonal, conforme pode ser observado na Figura 8.

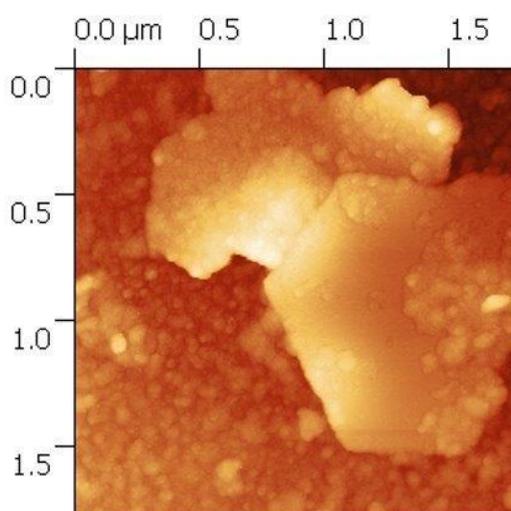


Figura 8 - Exemplo de uma placa hexagonal de HDL obtida por MFA para uma amostra sintetizada por co-precipitação nas seguintes condições: $pH = 10$, $T = 373$ K, $t = 4$ h. Retirado de [1].

1.7 - Aprendizagem automática em Ciência e Engenharia dos Materiais

A aprendizagem automática consiste numa ferramenta que explora um conjunto de dados que, posteriormente, constrói modelos para prever resultados e resolver problemas. Com o aparecimento e desenvolvimento da AA, esta tem vindo a permitir obter resultados mais rápidos e precisos, aprimorando a sua eficiência a nível temporal e económico para além de permitir uma melhor compreensão nas análises de resultados. Tal facto deve-se ao progresso que os cientistas têm vindo a desenvolver ao longo dos anos, como o aumento da velocidade e da capacidade de armazenamento de dados

computacionais, o que permite uma capacidade de recolha e de análise de um maior número de dados.

A área da ciência e engenharia dos materiais depende, fortemente, da investigação, de forma de conseguir analisar e compreender o comportamento dos materiais.

Além do mais, na ciência dos materiais, existe uma relação complexa entre a estrutura e as propriedades de interesse num material, o que poderá dificultar o estudo destas relações através de métodos de correlação lineares. Neste tipo de situações, os cientistas e engenheiros podem recorrer a métodos de AA para investigar a ciência e tecnologia dos materiais e otimizá-las, a partir de dados existentes na literatura [29].

Estes métodos são usados para prever os resultados e não, necessariamente, para a sua interpretação, porque, a partir de um conjunto de dados, é possível criar um modelo composto por vários algoritmos de AA, para obter previsões mais exatas [30], mas, uma vez por outra, de difícil interpretação.

Os modelos mais complexos exigem mais parâmetros de ajuste, o que torna as estimativas dos resultados mais complicadas para as previsões individuais poderem ser interpretadas. Todavia, os parâmetros de um modelo linear são mais simples e interpretáveis, mas podem não apresentar exatidão tão elevada como modelos mais complexos [31].

Assim sendo, a AA ajuda a prever o comportamento dos materiais, em diferentes ambientes, recorrendo a um processo mais eficiente do que através da tentativa e erro dos testes experimentais.

Aprendizagem Automática em HDL

Em 2017, os cientistas Hu *et al.* [32], criaram um modelo de Relação Quantitativa entre Estrutura e Propriedade (RQEP) para o espaçamento basal de HDL usando um algoritmo genérico de seleção de descritores e, seguidamente, foi efetuada regressão entre as propriedades e o espaçamento basal baseada no algoritmo máquina vetorial de relevância (MVR). O MVR foi proposto pela primeira vez por Tipping [33] em 2001 e foi aplicado com sucesso no campo de AA, como um modelo probabilístico disperso baseado em amostras limitadas. Um modelo RQEP liga as propriedades físicas, químicas e/ou físico-químicas de um composto à sua estrutura. Com estas informações este modelo tenta encontrar uma relação matemática entre a propriedade de interesse e os parâmetros descritivos (descritores). O modelo RQEP obtém um excelente desempenho, se empregar descritores e abordagem de modelação apropriados, sendo que estas duas etapas são cruciais para o resultado final. O erro relativo do modelo desenvolvido foi de 0,78% recorrendo a validação cruzada. Posteriormente, o modelo RQEP foi utilizado para recomendar um HDL com espaçamento basal desejado para síntese. O HDL sintetizado cumpriu o requisito do projeto destes cientistas, confirmando, deste modo, o poder de previsão do modelo desenvolvido [34].

Noutro trabalho realizado pelo mesmo grupo de investigação, este utilizou o valor da área superficial específica (ASE) do HDL para poder ser utilizado como o principal parâmetro de avaliação do adsorvente. Os membros desse grupo de investigação, utilizaram métodos de prospeção de dados para explorar as correlações da ASE (variando de 10 a 90 m²g⁻¹) com as suas composições químicas e parâmetros técnicos em busca de determinado material de HDL com ASE desejado. Naquela ocasião, utilizaram um método baseado no algoritmo genético em conjunto com regressão de suporte vetorial (RSV) para filtrar os principais descritores moleculares para modelação. O melhor modelo obtido foi, na ocasião, aplicado em amostras independentes para testar a capacidade de previsão no qual atingiram um erro médio relativo de 14,7% [32].

Em 2021, num trabalho conjunto realizado por Kailiang Lu *et al.*, estes autores consideraram que um dos problemas atuais é a proteção ambiental e a escassez de energia. Algumas soluções para os dispositivos de armazenamento de energia incluem condensadores convencionais, células de combustível, baterias recarregáveis e supercondensadores. Estes são os novos componentes que armazenam energia através de uma interface de duas camadas entre um eletrodo e um eletrólito [35]. Os

supercondensadores com hidróxidos duplos em camadas (HDL) têm capacitância específica excelente e um tempo de vida cíclico muito longo devido às suas estruturas únicas em camadas e riqueza em locais REDOX. O espaçamento basal (d) de HDL pode ser controlado selecionando cátions metálicos ideais e ânions interlamelares. De um modo geral, quanto maior o espaçamento d dos materiais dos elétrodos dos HDL, maior a capacitância específica dos supercondensadores. Nesse trabalho, o modelo de AA foi utilizado para procurar novos materiais HDL com maior espaçamento d . As abordagens combinadas de AA e do algoritmo genético foram utilizadas para selecionar o subconjunto de descritores apropriado, incluindo, parâmetros atômicos e composições químicas de HDL. O método *Extreme Gradient Boosting* foi usado para prever o espaçamento basal de HDL [35]

1.8 - Inovações do trabalho em relação ao estado da arte

Esta tese de mestrado tem como principal objetivo e inovação em relação ao estado da arte, conseguir prever se um determinado composto pode ser intercalado em HDL. Para atingir este objetivo, vários objetivos técnico-científicos mais específicos são propostos:

- 1) Obter e organizar o maior número de dados possível durante o tempo estabelecido, considerando alguns trabalhos de literatura e dados obtidos experimentalmente, tanto neste trabalho, como outros obtidos na UA e na SMT. Desta forma, quantos mais dados forem obtidos, possivelmente mais robustos serão os modelos desenvolvidos;
- 2) Com o conjunto de dados descritos em 1), obter um modelo que considera diferentes tipos de condições experimentais;
- 3) Utilizar uma linguagem de programação de código aberto, sem ser necessário recorrer a programas comerciais, o que possibilita a construção de aplicações em trabalhos futuros.
- 4) Realizar trabalho experimental em ambiente industrial, nomeadamente para obter HDL com diferentes espécies intercaladas, testar condições

experimentais e obter modelos mais ligados à produção industrial e à aplicação real dos materiais.

Capítulo 2 - Detalhes experimentais

2.1 - Materiais usados

Para a realização da síntese de HDL por co-precipitação (ZnAl-NO_3), foram utilizados os seguintes reagentes: hidróxido de sódio (NaOH , $\geq 99,0\%$, F. Mesquita Araújo, Lda.), nitrato de zinco hexahidratado ($\text{Zn(NO}_3)_2 \cdot 6\text{H}_2\text{O}$, $\geq 98\%$, F. Mesquita Araújo, Lda.) e nitrato de alumínio nonahidratado ($\text{Al(NO}_3)_3 \cdot 9\text{H}_2\text{O}$, $\geq 98\%$, Cfm Oskár Tropitzch, GmbH).

Foram utilizados 53 compostos, escolhidos aleatoriamente em ambos laboratórios da SMT e do DEMAC, para testar a sua intercalação nos HDL:

Tabela 1 - Compostos usados na intercalação em HDL.

Nome	Fornecedor	CAS	Massa Molecular / g mol^{-1}
Fluoresceína	Sigma-Aldrich	518-47-8	376,3
Rodamina B	Sigma-Aldrich	81-88-9	393,9
Ácido Cítrico	Sigma-Aldrich	77-92-9	192,1
Fenolftaleína	Sigma-Aldrich	77-09-8	318,3
Eosina Y	Sigma-Aldrich	15086-94-9	647,8
Lactona de cristal violeta	Sigma-Aldrich	1552-42-7	415,5
Sal de sódio de azul de bromotímol	Sigma-Aldrich	34722-90-2	646,4
Vermelho de cresol	Sigma-Aldrich	1733-12-6	382,4
Azul de tímolo	Sigma-Aldrich	76-61-9	466,6
Preto de eriocromo T	Alfa Aesar	1787-61-7	461,3
Verde de bromocresol	ACS	76-60-8	698,0
Rodamina 6G	Alfa Aesar	989-38-8	479,0
Sal de sódio de ibuprofeno	Sigma-Aldrich	31121-93-4	228,2
7-hidroxycumarina	Alfa Aesar	93-35-6	162,1
Cumarina	Sigma-Aldrich	91-64-5	146,1
Benzoato de sódio	VWR Chemicals	532-32-1	144,1
Ácido tricloroacético	Merck	76-03-9	163,3
Ácido fumárico	Alfa Aesar	110-17-8	116,0
Ácido ascórbico	VWR CHEMICALS	50-81-7	176,1

Ácido tânico	Sigma-Aldrich	1401-55-4	170,2
Ácido oxálico	Alfa Aesar	144-62-7	90,0
Benzotriazol	Chemical point	95-14-7	119,1
Ácido dietilentriaminopentaacético	Sigma-Aldrich	67-43-6	393,9
5-cloro-1,2,3-benzotriazol	Sigma-Aldrich	94-97-3	153,6
Sal de sódio do ácido glucónico	Sigma-Aldrich	526-95-4	218,1
Sal de sódio do DL-Malato	TCI	676-46-0	178,0
Sal de sódio do EDTA	Sigma-Aldrich	6381-92-6	372,2
3-amino-5-mercapto-1,2,4-triazol	TCI	16691-43-3	116,1
Dodecil sulfato de sódio	Acros	151-41-7	288,3
DL- Metionina	TCI	63-68-3	149,2
Ácido sulfâmico	VWR CHEMICALS	5329-14-6	97,0
2-mercaptopirimidina	Thermo Scientific	1450-85-7	112,1
Piridina-2-tiol	Apollo Scientific	2637-34-5	116,1
Ácido salicílico	Novacyl	69-72-7	138,1
Rosa bengala	Acros Organics	632-69-9	1017,6
Ácido esteárico	Peter Greven	822-16-2	306,4
Ácido 4-tiolbenzoico	Apollo Scientific	1074-36-8	154,1
1,2,4-triazol	Sigma-Aldrich	288-88-0	69,0
Benzimidazol	Sigma-Aldrich	51-17-2	118,1
Ácido 2,6-piridinadicarboxílico	Sigma-Aldrich	499-83-2	167,1
Acetato de sódio	Sigma-Aldrich	127-09-3	82,0
4-acetamidofenol	Acros Organics	103-90-2	151,1
2-aminofenol	Sigma-Aldrich	95-55-6	109,1
Cisteína	Acros Organics	52-90-4	121,1
Ácido maleico	Sigma-Aldrich	110-16-7	116,0
Salicilaldeído	Sigma-Aldrich	90-02-8	122,1
Curcumina	Sigma	458-37-7	368,3
4-aminofenol	Sigma-Aldrich	123-30-8	109,1
Sesamol	Sigma-Aldrich	533-31-3	138,1
Vermelho de metilo	Sigma-Aldrich	493-52-7	269,3
Naringenina	Sigma-Aldrich	10236-47-2	580,5
8-hidroxiquinolina	Sigma-Aldrich	148-24-3	145,1

Quercetina	Sigma-Aldrich	117-39-5	302,2
Pectina	Sigma	9000-69-5	150,1
EDTA	Sigma	60-00-4	292,2
Cafeína	Sigma-Aldrich	58-08-2	194,1
Alantoína	Sigma-Aldrich	97-59-6	158,1

EDTA = Ácido etilenodiaminotetracético

2.2 - Síntese do material precursor (HDL)

O material precursor ($Zn(2)Al-NO_3 - HDL$), usado depois para se substituir o anião nitrato pelos aniões pretendidos, foi preparado através do método de co-precipitação. A síntese foi realizada em atmosfera de azoto e todas as soluções foram preparadas com água desionizada, para evitar a contaminação com aniões carbonato. A proporção Zn:Al escolhida para a síntese foi de 2:1 para obter camadas estáveis de HDL.

Na primeira etapa, uma solução 0,5 M $Zn(NO_3)_2 \times 6H_2O$ e 0,25 M $Al(NO_3)_3 \times 9H_2O$ foi adicionada lentamente a 1,5 M $NaNO_3$, sob agitação vigorosa à temperatura ambiente (a adição levou aproximadamente 1,5 h). Durante a reação, o pH foi mantido constante ($pH = 10 \pm 0,5$) através da adição simultânea de NaOH 2 M. Em seguida, a pasta obtida foi submetida a um tratamento térmico a 100 °C durante 4 h para cristalização dos HDL. Após a síntese, o material foi lavado para retirar o excesso de sais.

Após a obtenção do material, este foi submetido a um sistema de filtração por vácuo utilizando quitasato, de maneira que o HDL fique na forma de pasta para poder ser intercalado com os materiais supra indicados. Seguidamente, foi realizada uma secagem a uma parte deste material, para que se determinasse a percentagem de sólido (20%).

2.3 - Intercalação dos compostos em HDL

O processo de intercalação dos compostos selecionados nos HDL iniciou-se com a determinação da massa necessária de composto em pó para 50 ml de água desionizada (0.1 M), assim como a quantidade de HDL necessária para a intercalação. De seguida, foi então pesado o composto e dissolvido em 50 ml de água desionizada, num copo de plástico utilizando um agitador, durante 30 a 45 min. Após este procedimento foi medido o pH (Tabela 2) e acertado consoante o pK_a do composto, de forma a que a molécula ficasse desprotonada, ou seja, o pH foi acertado de modo a ficar acima pelo menos do valor mais baixo de pK_a , utilizando uma solução básica de NaOH 2M e/ou uma solução

de ácido de HNO₃ 0.5M. Com isto, o HDL é disperso na solução e a mistura deixada a intercalar numa placa com agitadores magnéticos durante 24 h.

Tabela 2 - Valores de pKa dos compostos e pH das soluções preparadas.

Nome	pKa [36–38]	pH	Funcionalidade/Aplicação
Fluoresceína	6,7	8,0	Corante
Rodamina B	4,2	10,8	Corante
Ácido Cítrico	2,8	8,5	Antioxidante
Fenolftaleína	9,7	10,8	Indicador de pH
Eosina Y	2,0	8,3	Corante
Lactona de cristal violeta	4,7	10,9	Corante
Vermelho de cresol	8,3	9,0	Corante
Azul de timol	1,7	9,1	Indicador de pH
Preto de eriocromo T	4,9	8,4	Medidor da dureza da água
Verde de bromocresol	4,0	8,5	Corante indicador
Rodamina 6G	6,1	9,2	Corante
Sal de sódio de ibuprofeno	4,9	8,3	Fármaco
7-hidroxycumarina	7,7	8,8	Fármaco
Cumarina	-6,9	8,5	Fármaco/Industria
Benzoato de sódio	4,1	7,6	Antifúngico
Acido tricloroacético	0,5	8,5	Pomadas
Acido fumárico	3,0	8,0	Aditivo Alimentar
Acido ascórbico	4,7	8,7;10,0	Antioxidante
Acido tânico	6,0	8,0	Fármacos
Acido oxálico	1,4;4,4	9,4;8,0	Química analítica
Benzotriazol	8,3	8,1	Detergentes
Ácido dietilentriaminopentaoacético	1,8	8,1	Produtos químicos
5-cloro-1,2,3-benzotriazole	7,4	8,0;10,0;1 1,5	Indústria química
Sal de sódio do ácido glucónico	3,7	6,82;8,36; 10,3	Produtos químicos
Sal de sódio DL-Malato	3,5;5,0	6,8	Regulador de cidez/ industria alimentar
Sal de sódio EDTA	2,0; 2,7; 6,2; 10,3	4,2	Indústria química
3-amino-5-mercapto-1,2,4-triazol	12,5	8,3; 10,0	Inibidor de corrosão

Dodecil Sulfato de sódio	-1,5	10,1	Inibidor de corrosão / detergentes
DL- Metionina	2,2; 9,2	9,8; 8,0	Fármacos / indústria alimentar
Ácido sulfâmico	1,0	8,0; 10,2	Produtos de limpeza industrial
2-mercaptopirimidina	7,1	8,0	Catalisadores
Piridina-2-tiol	-1,0; 10,0	8,4; 10,2	Produtos cosméticos
Ácido salicílico	2,9	8,2; 9,4	Produtos estéticos
Rosa Bengala	3,9; 4,7	8,6	Corante
Ácido esteárico	4,9	10,8	Indústria têxtil
Ácido 4-tiolbenzoico	4,0	8,0; 8,9	Química orgânica
1,2,4-triazole	2,2	8,2; 10,0	Químicos agrícolas
Benzimidazol	5,5	8,0	Fungicidas
Ácido 2,6-piridinacarboxílico	3,2	8,3	Agente microbateria
Acetato de sódio	5,1	8,0	Indicador de pH
4-acetamidofenol	9,8	10,1	Fármacos
2-aminofenol	4,7; 9,9	10,0	Corantes
Cisteína	1,7	8,1	Antioxidante
Ácido maleico	1,8	8,4	Indústria plástica/ indústria química
Salicilaldeído	8,3	8,1	Agroquímicos / fármacos
Curcumina	8,0	10,1	Aplicações terapêuticas
4-aminofenol	5,4; 10,3	10,5	Indústria química / biossensores
Sesamol	10,0	10,0	Antioxidante
Vermelho de metilo	5,0	8,0	Corante indicador de pH
Naringenina	8,6	10,2	Antioxidante / anti- inflamatório
8-hidroxiquinolina	9,9	10,8	Fungicidas
Quercetina	6,4	8,7	Anti-inflamatórios
Pectina	3,5	8,3	Indústria alimentar
EDTA	0,2	8,0; 10,0	Indústria química
Cafeína	0,6	10,0	Indústria alimentar
Alantoína	8,5	10,0	Produtos cosméticos / fármacos

2.4 - Lavagem e secagem

Após as 24 h de intercalação, os compostos foram divididos em tubos de plástico de 15 ml e foram postos no equipamento de centrifugar (HERMLE, Z 446) (Figura 9)

para completar a primeira de 4 lavagens. De seguida, foi retirado o líquido em excesso para se poder adicionar água desionizada e misturada para que o material durante a lavagem fique o mais puro possível. Este processo foi repetido em todas as lavagens. Com este procedimento concluído, os compostos foram transportados para um forno para ser efetuada uma secagem a 60° C, com duração entre 24 a 48 horas.



Figura 9 - Equipamento de centrifugação

2.5 - Análise por difração raios-X

Após a secagem, analisou-se os compostos obtidos por DRX. Para isso, o composto foi moído com a ajuda do almofariz e de um pistilo. A seguir, com ajuda de uma espátula, carregou-se um porta-amostras com uma pequena porção do material a analisar no equipamento (*Rigaku*, modelo - *MiniFlex600-C*). Após terminada a análise, todo o material de análise, como o porta amostras, espátula, almofariz e pistilo, zona de análise do equipamento, era submetido a uma limpeza com etanol para se efetuar novas análises e, assim, evitar contaminações.

Capítulo 3 - Aprendizagem automática

3.1 - Passos para a criação e utilização do modelo AA

A linguagem de programação Python foi usada para desenvolver um modelo de aprendizagem automática. Esta linguagem permite a manipulação, análise e visualização de dados, de forma estruturada, aberta e gratuita, sendo amplamente usada em diversas áreas de investigação.

Os dados obtidos experimentalmente, foram organizados sob a forma de uma base de dados manipulada em Excel. Seguidamente, elaborou-se um código Python denominado de “LDH.ipynb”, para a sua exploração, análise estatística, e desenvolvimento de um modelo preditivo de aprendizagem automática.

O trabalho foi efetuado recorrendo à plataforma Google Colab, que é um serviço de nuvem gratuito hospedado pelo própria Google, para incentivar a investigação recorrendo a aprendizagem automática e inteligência artificial. Nesta plataforma foi inserida a base de dados completa onde posteriormente foi necessário realizar vários passos de código para criar o modelo de AA.

Desta forma, abaixo serão descritos os pontos necessários para construir o modelo:

- 1) Base de dados
- 2) Tratamento de dados
- 3) Descritores
- 4) Seleção de descritores
- 5) Florestas aleatórias e sua otimização
- 6) Validação e teste do modelo
- 7) Importância dos descritores pelo método de permuta

3.1.1 - Base de dados

A base de dados é composta por 98 linhas e 40 colunas perfazendo assim um total de 3920 células. Nesta base de dados, para além dos resultados das sínteses do estágio (71 sínteses) também foram inseridos resultados de sínteses efetuadas anteriormente na UA (17 sínteses) e na SMT (9 sínteses), como mostra a Figura 10.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
1	M2+	M3+	Ratio	Anion_initial	Compound_intercalate	CAS	H2O	Methodology	pH	Time_h	n_times	Temperature_K	C_Anion_M	Washing_n_times	Subs
2	Zn	Al	2	Nitrate	2-Mercaptobenzothiazole	149-30-4	-	Anion-exchange	10	24	2	Room temperature	0.1	4	Yes
3	Zn	Al	2	Nitrate	Picolinic acid	98-98-6	-	Anion-exchange	8	24	2	Room temperature	0.1	4	Yes
4	Zn	Al	2	Nitrate	Picolinic acid	98-98-6	-	Anion-exchange	6	24	2	Room temperature	0.1	4	Yes
5	Zn	Al	2	Nitrate	Nicotinic acid	59-67-6	-	Anion-exchange	6	24	2	Room temperature	0.1	4	Yes
6	Zn	Al	2	Nitrate	Nicotinic acid	59-67-6	-	Anion-exchange	8	24	2	Room temperature	0.1	4	Yes
7	Zn	Al	2	Nitrate	Salicylic acid	69-72-7	-	Anion-exchange	6	24	2	Room temperature	0.1	4	Yes
8	Zn	Al	2	Nitrate	Salicylic acid	69-72-7	-	Anion-exchange	8	24	2	Room temperature	0.1	4	Yes
9	Zn	Al	2	Nitrate	Thiolatic acid	79-42-5	-	Anion-exchange	8	24	2	Room temperature	0.1	4	Yes
10	Zn	Al	2	Nitrate	Thionicotinic acid	51087-03-7	-	Anion-exchange	8	24	2	Room temperature	0.1	4	Yes
11	Zn	Al	2	Nitrate	Thionicotinic acid	51087-03-7	-	Anion-exchange	6	24	2	Room temperature	0.1	0	Yes
12	Zn	Al	2	Nitrate	Thionicotinic acid	51087-03-7	-	Anion-exchange	6	24	2	Room temperature	0.1	4	Yes
13	Zn	Al	2	Nitrate	Isonicotinic acid	55-22-1	-	Anion-exchange	6	24	2	Room temperature	0.1	0	Yes
14	Zn	Al	2	Nitrate	Isonicotinic acid	55-22-1	-	Anion-exchange	6	24	2	Room temperature	0.1	4	Yes
15	Zn	Al	2	Nitrate	Isonicotinic acid	55-22-1	-	Anion-exchange	8	24	2	Room temperature	0.1	4	Yes
16	Zn	Al	2	Nitrate	Thiosalicylic acid	147-93-3	-	Anion-exchange	6	24	2	Room temperature	0.1	0	Yes
17	Zn	Al	2	Nitrate	Thiosalicylic acid	147-93-3	-	Anion-exchange	6	24	2	Room temperature	0.1	4	Yes
18	Zn	Al	2	Nitrate	Thiosalicylic acid	147-93-3	-	Anion-exchange	8	24	2	Room temperature	0.1	4	Yes
19	Zn	Al	2	Nitrate	5-Chlorobenzotriazole	94-97-3	-	Anion-exchange	8	24	2	Room temperature	0.1	4	No
20	Zn	Al	2	Nitrate	Citric acid sodium salt	6132-04-3	-	Anion-exchange	8	24	2	Room temperature	0.1	4	Yes
21	Zn	Al	2	Nitrate	D-Gluconic acid sodium salt	527-07-1	-	Anion-exchange	8	24	1	Room temperature	0.1	4	Yes
22	Zn	Al	2	Nitrate	Nitrate	14797-55-8	-	Anion-exchange	8	24	1	Room temperature	0.1	4	Yes
23	Zn	Al	2	Nitrate	Nitrite	14797-65-0	-	Anion-exchange	8	24	1	Room temperature	0.1	4	Yes
24	Zn	Al	2	Nitrate	1,2,3-Benzotriazole	95-14-7	-	Anion-exchange	10	24	1	Room temperature	0.1	4	Inco
25	Zn	Al	2	Nitrate	2-Mercaptobenzothiazole	149-30-4	-	Anion-exchange	11	24	1	Room temperature	0.1	4	Yes
26	Zn	Al	2	Nitrate	Ascorbic acid	50-81-7	-	Anion-exchange	8	24	1	Room temperature	0.1	4	Yes
27	Zn	Al	2	Nitrate	4-Aminobenzoic acid	150-13-0	-	Anion-exchange	8	24	1	Room temperature	0.1	4	Yes

Figura 10 - Imagem parcial da base de dados.

3.1.2 - Tratamento de dados

Após obter a base de dados, seguiu-se o tratamento de dados, que foi efetuado recorrendo à livraria *Pandas* [39]. Deste modo, foram analisados vários pontos tais como:

- Possíveis linhas repetidas;
- Contagem do número de compostos;
- Erros na escrita de palavras
- Valores em falta;
- Estimar valores para o descritor *XLogP3* (ver definição no ponto 3.1.3), nos casos em que não foi possível ser calculado (apenas 2 compostos), tentando-se apurar a sua correlação com as restantes propriedades, e, como isto não aconteceu, substitui-se, esses dois exemplos, pelo seu valor médio.

3.1.3 - Seleção dos descritores

Para este estudo, foram considerados 17 descritores para cada composto, determinados pelas condições da síntese e também através das propriedades dos compostos. (53 compostos).

Dois deles são descritores associados à própria síntese:

- pH;
- número de passos de intercalação.

Os restantes estão associados às propriedades das moléculas em si, e foram recolhidos a partir de diversas fontes [36–38]:

- temperatura de fusão;
- pKa;
- peso molecular;
- *XLogP3*;
- número de doadores/aceitadores de ligações de hidrogénio;
- número de ligações rotativas;
- área de superfície topológica;
- número de átomos pesados;
- complexidade;

- número de estereocentros de átomos definidos/indefinidos;
- número de estereocentros através de ligação definida;
- número de unidades ligadas covalentemente;
- subtração do pKa ao pH.

O **pH** é uma escala numérica adimensional utilizada para especificar a acidez ou basicidade de uma solução aquosa.

O **número de passos de intercalação** é a quantidade de vezes a que o composto foi submetido a sínteses.

A **Temperatura de fusão** designa a temperatura à qual uma substância passa do estado sólido ao estado líquido.

O **pKa**, a constante de acidez K_a , também chamada constante de dissociação ácida ou constante de ionização ácida, é uma constante de equilíbrio que exprime o grau de dissociação para um dado ácido de Bronsted numa reação de equilíbrio químico. Aquela constante possui definições diferentes para ácidos fortes, moderados e fracos.

O **Peso molecular**, refere-se à soma das massas dos átomos da molécula.

O **XLogP3** é um método para o cálculo rápido de logP, este é o coeficiente de partição é definido como a relação das concentrações da substância em um solvente orgânico e em água, dado que ambos os solventes são necessariamente imiscíveis. Métodos mais antigos como XLOGP e XLOGP2, que são métodos aditivos atômicos com fatores de correção bem definidos, e já ganharam popularidade neste campo pela sua aplicabilidade e razoável exatidão.

O **Número de átomos doadores/aceitadores de ligações de hidrogénio** está ligado às propriedades do doador e recetor de ligações de hidrogénio dos átomos na molécula.

O **Número de ligações rotativa** representa o número de ligações que permitem rotação livre em torno de si. Estas são definidas como qualquer ligação simples, não num anel, ligadas a um átomo pesado não terminal. Excluídas da contagem estão as ligações amida C–N devido à sua elevada barreira de energia rotacional.

A **Área de superfície topológica** é uma medida pertinente da área de superfície polar que evita a necessidade de calcular a estrutura 3D do ligante ou decidir qual é a conformação ou conformações biológicas relevantes.

O **Número de átomos pesados** é o número total de átomos diferentes de hidrogénio (átomo 'pesado') dentro da estrutura química.

A **Complexidade** é uma classificação dos múltiplos aspetos de cada composto em que é calculada a estimativa aproximada do quão complexa é a estrutura, do ponto de vista dos elementos contidos e das características estruturais exibidas, incluindo a simetria.

O **Número de estereocentros de átomos definidos/indefinidos** é determinado por um estereocentro que tem quatro grupos ligados a si que são únicos, de modo que, se quaisquer dois grupos ligados entre si mudarem a sua orientação, será criada uma nova molécula.

O **Número de unidades ligadas covalentemente** é uma ligação química que envolve a partilha de eletrões para formar pares de eletrões entre os átomos. Esses pares de eletrões são conhecidos como pares compartilhados ou pares de ligação.

O **Número de estereocentros através de ligação definida** é marcada como uma localização (ponto) dentro de uma molécula, em vez de um átomo específico, no qual a troca de dois grupos cria um estereoisómero.

A **subtração do pKa ao pH** é importante para estimar o grau de desprotonação da molécula. Quanto maior for o valor do pH em relação ao pKa, maior o grau de desprotonação da molécula.

Estas propriedades foram calculadas para as moléculas na sua forma neutra. Foi uma aproximação, uma vez que era mais fácil aceder a propriedades calculadas para a forma neutra e, por outro lado, em vários casos, é complicado saber exatamente qual a carga formal da molécula.

Para selecionar os descritores mais importantes para serem incluídos no modelo final de aprendizagem automática, estes foram selecionados usando a eliminação recursiva de descritores (ERD).

A abordagem de ERD foi realizada com validação cruzada integrada e o algoritmo de florestas aleatórias. Somente oitenta por cento de todo o conjunto de dados previamente subdividido para treino/validação foi usado para seleção de descritores. O tratamento de ERD seleciona os descritores do modelo considerando recursivamente um número cada vez menor de descritores em cada etapa, conforme representado na Figura 11.

Primeiro, o modelo foi treinado com o conjunto inicial contendo todos os descritores, e a importância de cada recurso é considerada tendo em conta uma determinada propriedade estatística ou métrica de desempenho (neste caso, foi considerado o coeficiente de correlação de Matthews [CMM]).

Em seguida, em cada etapa, é eliminada a característica menos importante do conjunto de descritores a serem considerados na próxima etapa.

Este procedimento foi repetido recursivamente para cada subconjunto de descritores até que apenas um descritor fosse obtido. O conjunto de descritores selecionado foi considerado aquele que possuía o valor mais elevado em relação à métrica de desempenho do modelo selecionada. Uma avaliação final dos descritores selecionados foi realizada com o modelo de florestas aleatórias desenvolvido, comparando as métricas de desempenho de um modelo com os descritores selecionados e com um modelo considerando todos os descritores iniciais disponíveis.

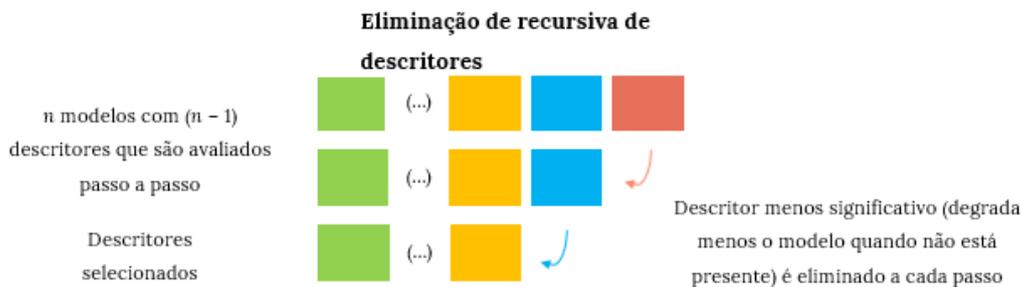


Figura 11 - Representação do modelo de eliminação recursiva de descritores. Adaptado [40].

3.1.4 - Métodos de AA: Otimização do modelo

Os métodos de AA classificaram as parcelas quanto aos seus descritores, obtendo um modelo preditivo. Para este estudo, o método usado foi o das **florestas aleatórias**, uma vez que é um método que apresenta bons resultados, exige poucos recursos computacionais e é relativamente fácil de obter os melhores parâmetros.

Na construção do modelo, este algoritmo é uma combinação de um conjunto de árvores de decisão com o método de agregação, obtendo uma previsão mais exata e estável.

O método de árvores de decisão divide os dados em parcelas cada vez mais pequenas, indicando várias possibilidades de decisão até chegar à tomada de decisão final baseada em diversos descritores. Este método é parecido como uma árvore, como é possível analisar pela Figura 12, porque o conjunto de dados corresponde à raiz da árvore e estes vão sendo divididos conforme a distribuição dos seus ramos (nós de decisão), que, por sua vez, continuam a dividir-se em parcelas mais pequenas nas ramificações até

chegarem aos nós das folhas. Quando se chega a um determinado ponto, o nó final contém a mesma classe para quase todas as entradas de dados, ou quando não existem descritores para dividirem mais os dados, a árvore de decisão termina [41].

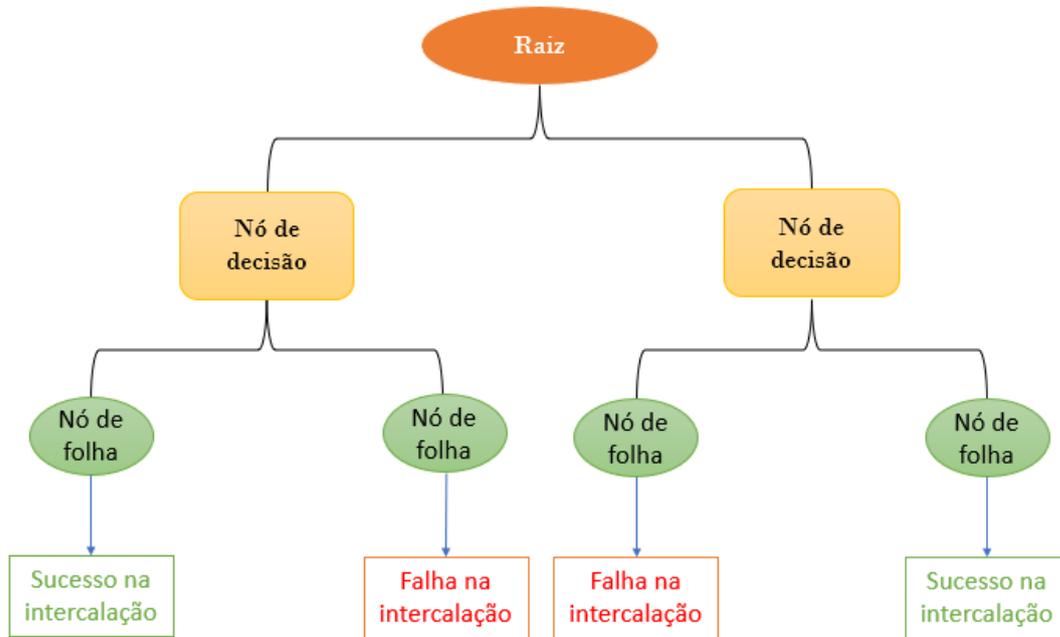


Figura 12 - Representação da execução do método de árvores de decisão.

O método de agregação utiliza vários subconjuntos do conjunto de dados, e, ao combiná-los, dá origem a vários modelos. A criação de múltiplos modelos permite obter melhores resultados do que apenas com um único modelo [41]. Contudo, estes modelos não podem ser totalmente independentes, pois, caso contrário, exigiria ter muitos dados.

Deste modo, são criadas várias amostras aleatórias (*bootstrapping*), para que cada uma atue como outro conjunto de dados quase independente. Posteriormente, estas amostras são agregadas para obter um modelo do conjunto com uma menor variação.

Uma vez que, neste estudo, se recorreu a métodos de classificação, calcula-se que a média das probabilidades de cada classe seja devolvida por todos os modelos, sendo o resultado final a classe mais vezes obtida por todos os modelos [42]. A Figura 13 representa a execução do modelo de agregação.

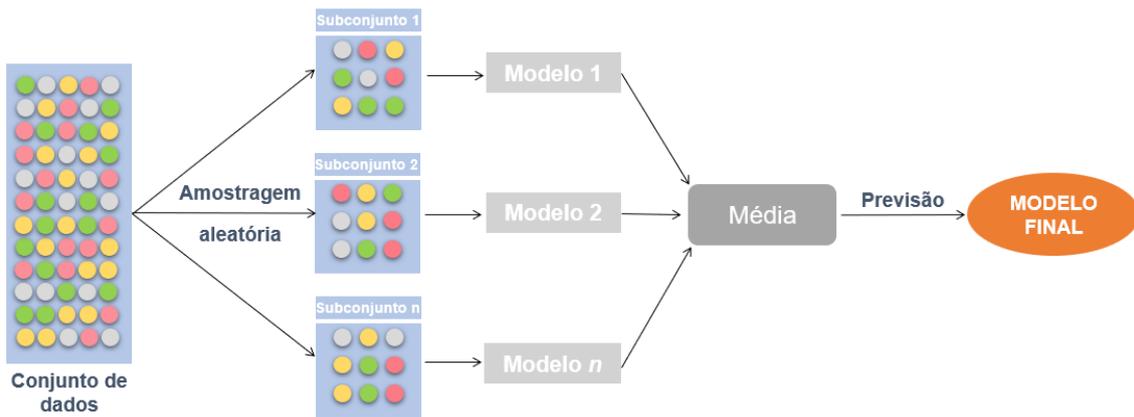


Figura 13 - Representação da execução do método de agregação. Retirado de [42].

Na Figura 14, está representado o método de florestas aleatórias no qual este combina a agregação das amostras com agregação das propriedades, considerando um número menor de propriedades em cada subconjunto para descrever o resultado final. Assim, é gerada uma combinação de árvores diferentes, recorrendo a uma amostragem aleatória para gerar subconjuntos de florestas aleatórias das amostras e propriedades.

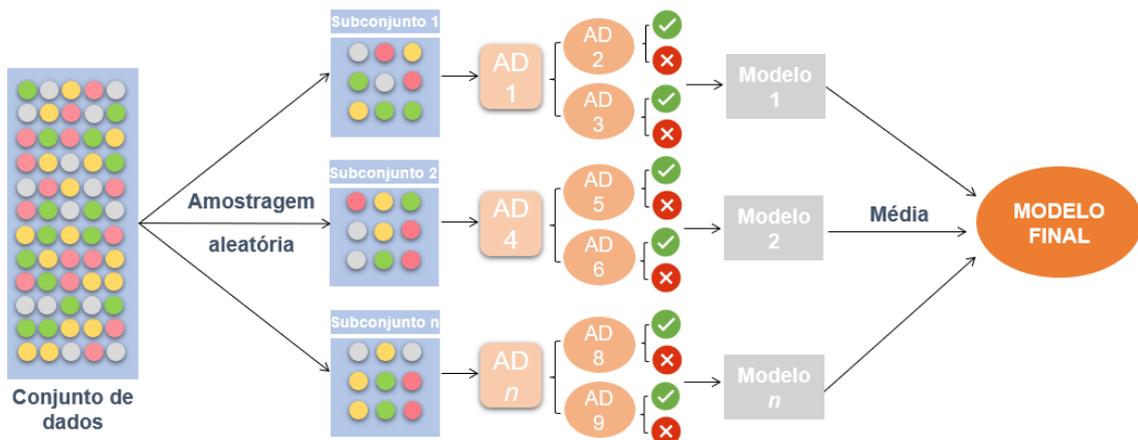


Figura 14 - Representação da execução do método das florestas aleatórias. Retirado de [42].

3.1.5 - Validação e teste do modelo, e métricas estatísticas

Para validar e testar o modelo, recorreu-se à validação cruzada e a um teste final. A validação cruzada é um método usado para avaliar o desempenho (ou exatidão) dos modelos de AA, e escolher os melhores descritores e parâmetros do modelo, sem ter uma tendência para um determinado grupo de dados.

Na validação cruzada, é criado um número fixo de subconjuntos (ou parcelas) dos dados, executando a validação em cada parcela (usando as restantes para treino). No final, calcula-se a média dos erros de todas as parcelas para a estimativa de erro geral. A validação cruzada serve para selecionar o melhor modelo, tendo em conta a otimização de todos os parâmetros do mesmo.

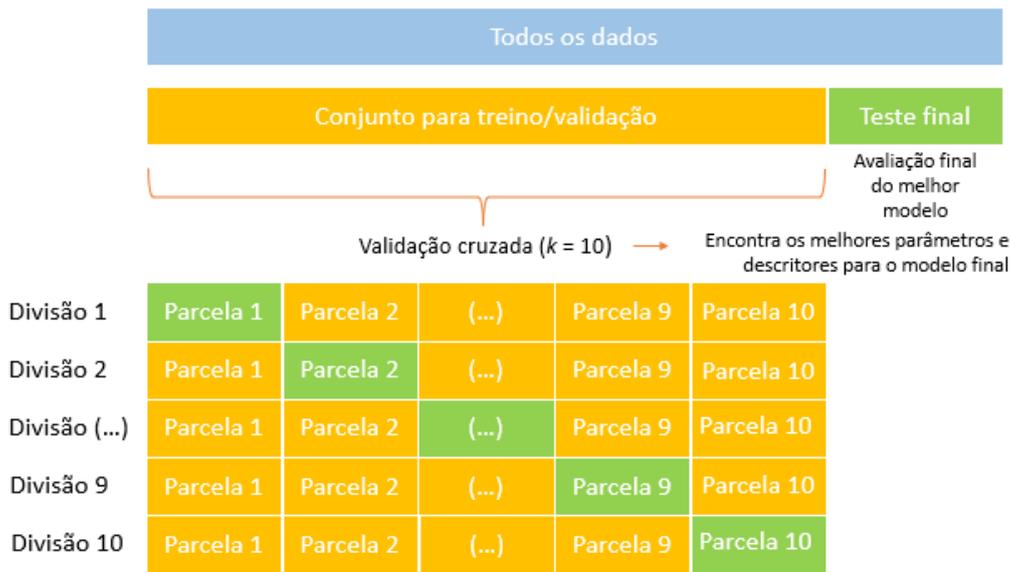


Figura 15 - Esquema de validação cruzada.

O teste serve para fazer uma avaliação final da capacidade de previsão do modelo otimizado para novos dados.

Vinte por cento de todo o conjunto de dados foi reservado como conjunto de teste final. Os exemplos considerados para o conjunto de dados de teste foram retirados antes de qualquer seleção de descritores ou otimização de modelo ser realizada. Os restantes oitenta por cento do conjunto de dados inicial foram usados para seleção de descritores e validação do modelo, empregando validação cruzada de 10 partes para encontrar a arquitetura e os hiperparâmetros ideais do modelo de florestas aleatórias.

A validação cruzada k -partes é um método que permite reduzir eventuais tendências do modelo para um determinado subconjunto de dados e, considerando $k = 10$, demonstrou ter um bom desempenho em relação à tendência e erro de previsão [42].

Para a avaliação deste método de AA são utilizadas várias métricas estatísticas, tais como a precisão, a exatidão, a sensibilidade, o valor de Kappa, e o coeficiente de correlação de Matthews.

Para este estudo, a classe para o sucesso da intercalação é a positiva e para o insucesso da intercalação é a negativa. A relação entre as duas classes pode originar quatro resultados: verdadeiros positivos (VP), classificados corretamente como sucesso da intercalação; verdadeiros negativos (VN), classificados corretamente como insucesso da intercalação; falsos positivos (FP), insucesso da intercalação classificado incorretamente como sucesso na intercalação; e falsos negativos (FN), sucesso na intercalação classificado incorretamente como insucesso na intercalação. Estes marcadores constituem uma matriz de confusão (Figura 16), na qual a maioria das métricas de AA para problemas de classificação são calculadas com base nesta matriz.

		VALORES PREVISTOS	
		SUCESSO	INSUCESSO
VALORES REAIS	SUCESSO	Verdadeiros positivos	Falsos Negativos
	INSUCESSO	Falsos Positivos	Verdadeiros Negativos

Figura 16 - Matriz de confusão, onde a classificação do sucesso da intercalação resulta em 4 possibilidades diferentes.

A exatidão (eq. 1) é uma das métricas mais usadas para a tarefa de AA de classificação, sendo que uma previsão só é considerada exata se os valores previstos estiverem próximos dos valores reais (aceite como referência), mas nem sempre é uma boa métrica de avaliação de uma matriz de confusão, quando as classes estão desequilibradas, ou seja, quando o número de ocorrências da classe positiva e negativa são bastante diferentes na base de dados. Assim, recorre-se à exatidão balanceada (eq. 2), que é calculada a partir da média da sensibilidade (eq. 3) e especificidade (eq.4), para dar um peso igual ao sucesso da intercalação e insucesso da intercalação quando se avalia a exatidão.

A sensibilidade mede a probabilidade dos verdadeiros sucessos da intercalação serem classificados corretamente, enquanto a especificidade refere-se à probabilidade da correta classificação dos casos de insucesso de intercalação.

A precisão (eq. 5) é a proporção entre os positivos corretamente identificados (verdadeiros positivos) e todos os positivos identificados.

$$\text{Exatidão} = \frac{VP + VN}{VP + FN + FP + VN} \quad (\text{eq. 1})$$

$$\text{Exatidão balanceada} = \frac{\text{Sensibilidade} + \text{Especificidade}}{2} \quad (\text{eq. 2})$$

$$\text{Sensibilidade} = \frac{VP}{VP + FN} \quad (\text{eq. 3})$$

$$\text{Especificidade} = \frac{VN}{VN + FP} \quad (\text{eq. 4})$$

$$\text{Precisão} = \frac{VP}{VP + FP} \quad (\text{eq. 5})$$

Também é importante avaliar a confiabilidade das próprias métricas de avaliação dos modelos de AA. Esta é um fator importante, pois pode ajudar a encontrar erros na avaliação e interpretação dos resultados. Assim, o valor de Kappa permite testar a confiabilidade das métricas estatísticas para avaliar o modelo. É um valor padronizado e pode variar entre -1 e 1, e pode haver valores abaixo de 0 que significa que o acordo entre as previsões e os valores reais é pior do que a atribuição de valores à sorte, onde 0 representa nenhum nível de confiança no modelo e 1 representa um nível de confiança perfeito. Na Tabela 3 - Interpretação dos valores de Kappa. Tabela 3 está representada a interpretação do valor de Kappa.

Tabela 3 - Interpretação dos valores de Kappa.

Valor de Kappa	Interpretação
≤ 0	Nenhum acordo
0,0 – 0,20	Algum acordo
0,21 – 0,40	Acordo moderado
0,41 – 0,60	Acordo substancial
0,61 – 0,80	Muito bom acordo
0,81 – 1,0	Quase perfeito

Em estatística, o coeficiente phi é uma medida de associação para duas variáveis binárias (por exemplo, 0 ou 1; Sim ou Não). Em aprendizagem automática, este coeficiente é conhecido como coeficiente correlação de Matthews (CCM, eq.6) e é usado como medida da qualidade de classificações binárias. O coeficiente de correlação de Matthews, tal como o valor de Kappa, permite avaliar a confiabilidade do modelo, sendo a sua interpretação semelhante ao valor de Kappa.

$$CCM = \frac{(VN \times VP) - (FN \times FP)}{\sqrt{(VP + FP)(VP + FN)(VN + FP)(VN + FN)}} \quad (\text{eq. 6})$$

3.1.6 - Importância dos descritores pelo método de permuta

A importância de descritores pelo método de permuta é um algoritmo independente do modelo que elimina recursivamente descritores usando como figura de mérito a importância de permutação.

Esta medida avalia a importância de cada descritor, permutando aleatoriamente os seus valores e, seguidamente, avalia o seu efeito na diminuição no desempenho do modelo. A suposição é que um descritor importante, quando os seus valores são baralhados aleatoriamente, levará a uma queda significativa no desempenho do modelo, indicando sua importância. Esse algoritmo pode ser aplicado a qualquer modelo de aprendizagem automática e fornece uma medida imparcial de importância do descritor, que não depende do algoritmo selecionado.

Capítulo 4 - Apresentação e discussão de resultados

Neste capítulo, irão ser apresentados os resultados experimentais da síntese de HDL com compostos funcionais intercalados, assim como a interpretação dos resultados recorrendo a técnicas de aprendizagem automática.

4.1 - Resultados gerais das intercalações

Na Tabela 4 estão apresentados os resultados da intercalação nos HDL dos vários compostos testados, assim como as diversas condições usadas que foram variadas, tais como o pH e o número de intercalações, bem como o resultado final da intercalação (sucesso ou insucesso da intercalação, e o espaçamento basal para os casos em que a intercalação foi bem-sucedida). Neste trabalho foram também analisadas 26 sínteses anteriormente realizadas, nos laboratórios da UA e na SMT, tendo-se realizado neste estágio 71 sínteses, fazendo um total de 97 intercalações.

É possível observar que 50 das 97 intercalações foram claramente bem-sucedidas, ou seja, foi possível observar a intercalação do composto no HDL através de DRX, assim como o respetivo pico que permite estimar o espaçamento basal [$d(003)$]. Em 36 dos casos também foi claro que não houve sucesso na intercalação, restando assim 11 casos que deixaram mais dúvidas, estando referidos na tabela como intercalação parcial. Alguns destes casos irão ser analisados, com mais detalhe, mais adiante, tendo em conta os seus difratogramas, assim como dados da literatura. Para facilitar, ao longo da discussão os compostos serão referidos no texto com o nome da respetiva forma neutra, mas, na verdade, eles são intercalados dentro dos HDL na sua forma aniónica.

Tabela 4 - Principais condições de intercalação, resultados das diferentes intercalações, e distância entre camadas.

Nome	pH	Nº Intercalações	Sucesso	$d(003)$
Fluoresceína	8,0	1	Sim	16,23
Nitrato	8,0	1	Sim	8,85
Nitrato*	8,0	1	Sim	8,72
Rodamina B	8,7	1	Não	-
Ácido Cítrico	8,5	1	Sim	11,98
Fenolftaleína	10,8	1	Não	-
Eosina Y	8,3	1	Não	-

Lactona de cristal violeta	10,9	1	Não	-
Vermelho de cresol	9,0	1	Não	-
Azul de timol	9,1	1	Não	-
Preto de eriocromo T	8,4	1	Sim	10,84
Verde de bromocresol	8,5	1	Não	-
Rodamina 6G	9,2	1	Não	-
Sal de sódio de ibuprofeno	8,3	1	Sim	21,85
7-hidroxycumarina	8,8	1	Não	-
Cumarina	8,5	1	Não	-
Benzoato de sódio	7,6	1	Sim	15,46
Ácido tricloroacético	8,5	1	Sim	13,65
Ácido fumárico	8,0	1	Sim	11,70
Ácido ascórbico	8,7	1	Parcial	8,56
Ácido ascórbico	10,0	1	Parcial	8,20
Ácido ascórbico*	8,0	1	Parcial	8,93
Ácido tânico	8,0	1	Não	-
Ácido oxálico	8,0	1	Sim	8,19
Ácido oxálico	9,4	1	Sim	7,92
1,2,3-benzotriazol	8,1	1	Não	-
1,2,3-benzotriazol*	10,0	1	Não	-
Ácido dietilentriaminopentaacético	8,1	1	Não	-
5-cloro-1,2,3-benzotriazol	8,0	1	Não	-
5-cloro-1,2,3-benzotriazol	10,0	1	Não	-
5-cloro-1,2,3-benzotriazol	11,5	1	Não	-
5-cloro-1,2,3-benzotriazol*	8,0	1	Não	-
Sal de sódio do ácido glucónico	6,82	1	Sim	13,84
Sal de sódio do ácido glucónico	8,36	1	Sim	13,52
Sal de sódio do ácido glucónico	10,3	1	Sim	13,75
Sal de sódio do ácido glucónico*	8,0	1	Sim	13,57
Sal de sódio do DL-Malato	6,8	1	Sim	11,77
Sal de sódio do EDTA	4,2	1	Sim	13,92
3-amino-5-mercapto-1,2,4-triazol	8,3	1	Parcial	11,09
3-amino-5-mercapto-1,2,4-triazol	10,0	1	Não	-

Dodecil sulfato de sódio	10,1	1	Sim	26,35
DL- Metionina	8,0	1	Sim	16,05
DL- Metionina	9,3	1	Sim	19,35
Ácido sulfâmico	8,0	1	Parcial	8,30
Ácido sulfâmico	10,2	1	Não	-
2-mercaptopirimidina	8,0	1	Não	-
Piridina-2-thiol	8,4	1	Não	-
Piridina-2-thiol	10,2	1	Parcial	9,95
Ácido salicílico	8,2	1	Parcial	21,97
Ácido salicílico	9,4	1	Não	-
Ácido salicílico*	6,0	2	Sim	15,29
Ácido salicílico*	8,0	2	Sim	15,43
Rosa Bengala	8,6	1	Não	-
Ácido esteárico	10,8	1	Sim	24,8
Ácido 4-tiolbenzoico	8,0	1	Não	-
Ácido 4-tiolbenzoico	8,9	1	Sim	22,87
1,2,4-triazole	8,2	1	Parcial	8,19
1,2,4-triazole	10,0	1	Sim	8,31
Benzimidazol	8,0	1	Não	-
Ácido 2,6-piridinadicarboxílico	8,3	1	Não	-
Acetato de sódio	8,0	1	Sim	8,68
4-acetamidofenol	10,1	1	Não	-
2-aminofenol	10,0	1	Sim	12,89
Cisteína	8,1	1	Sim	15,79
Ácido maleico	8,4	1	Sim	7,59
Salicilaldeído	8,1	1	Parcial	16,85
Curcumina	10,1	1	Não	-
4-aminofenol	10,5	1	Não	-
Sesamol	10,0	1	Não	-
Vermelho de metilo	8,0	1	Sim	17,83
Naringenina	10,2	1	Sim	20,02
8-hidroxiquinolina	10,8	1	Parcial	12,59
Quercetina	8,7	1	Não	-

Pectina	8,3	1	Parcial	11,92
EDTA	8,0	1	Não	-
EDTA	10,0	1	Não	-
Cafeína	10,0	1	Não	-
Alantoína	10,0	1	Não	-
2-mercaptobenzotiazole*	10,0	2	Sim	16,67
2-mercaptobenzotiazole*	11,0	1	Sim	17,00
Ácido Picolínico*	6,0	2	Sim	21,44
Ácido Picolínico*	8,0	2	Sim	7,55
Ácido Nicotínico*	6,0	2	Sim	15,37
Ácido Nicotínico*	8,0	2	Sim	15,23
Ácido Tiolático*	8,0	2	Sim	11,03
Ácido Tionicotínico*	8,0	2	Sim	14,76
Ácido Tionicotínico*	6,0	2	Sim	15,22
Ácido Tionicotínico*	6,0	2	Sim	14,95
Ácido Isonicotínico*	8,0	2	Sim	14,69
Ácido Isonicotínico*	6,0	2	Sim	14,76
Ácido Isonicotínico*	6,0	2	Sim	14,89
Ácido Tiosalicílico*	8,0	2	Sim	14,95
Ácido Tiosalicílico*	6,0	2	Sim	13,35
Ácido Tiosalicílico*	6,0	2	Sim	13,57
Citrato de sódio di-hidratado*	8,0	1	Sim	11,07
Nitrito*	8,0	1	Sim	7,87
Ácido 4-Aminobenzoico*	8,0	1	Sim	15,02

* – Resultados obtidos anteriormente a este trabalho

Como exemplo de uma intercalação bem-sucedida pode referir-se a intercalação do ácido cítrico no HDL (HDL-citrato), cuja estrutura pode ser comparada com o HDL intercalado com nitrato (HDL-nitrato), uma vez que este é o material precursor. Através dos resultados de DRX representados no gráfico da Figura 17, foi possível observar que o difratograma do HDL-citrato representa uma estrutura monofásica, com picos bem definidos a 2θ iguais a $7,37^\circ$, $14,71^\circ$ e $22,01^\circ$. Estes picos correspondem às reflexões dos planos (003), (006) e (009), respetivamente. O HDL-nitrato também apresenta uma estrutura monofásica, onde os picos estão definidos em 2θ , a $9,98^\circ$, $19,98^\circ$ e em $30,15^\circ$,

cada um correspondente, respetivamente, à reflexão (003), (006) e (009). A estrutura do HDL do nitrato apresenta picos bem definidos e simétricos indicando que este HDL estava bem cristalizado. Também foi possível calcular o espaçamento basal HDL como sendo o $d(003)$ do nitrato igual a 8,85 Å e o $d(003)$ do citrato igual a 11,98 Å. Desta maneira, é possível afirmar que houve sucesso na troca aniónica uma vez que houve um aumento do espaçamento basal, dado que o citrato é uma molécula com uma dimensão superior ao nitrato. Além disso, a ausência de reflexões residuais no HDL-citrato que pudessem ser atribuídas ao HDL-nitrato garante a substituição completa dos aniões intercalados. Nas restantes moléculas intercaladas, cujo resultado foi bem-sucedido, registou-se um comportamento idêntico em relação às mudanças verificadas nos seus difratogramas em comparação com o difratograma do material precursor.

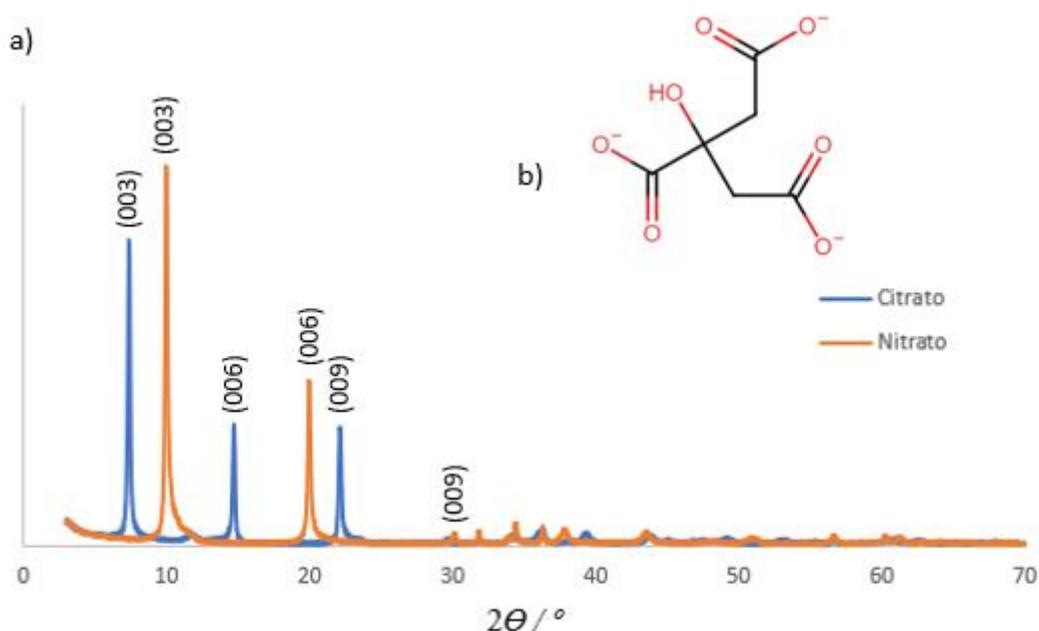


Figura 17 - DRX do HDL-NO₃ e do HDL-citrato; b) Molécula de citrato.

Um caso em que não foi possível observar a intercalação foi o do composto benzotriazole (BTA), quer recorrendo a um pH de síntese igual a 8,1 ou a 10,0. Os principais picos de difração deste composto estão representados na Figura 18 e na Figura 19, encontrando-se os picos do NO₃⁻ a $2\theta = 9,98^\circ$, $19,98^\circ$, e $30,15^\circ$. No difratograma do BTA (pH = 8,1) poderia, na melhor das hipóteses, ser considerado o pico correspondente à reflexão (003) a $2\theta = 10,20^\circ$, e o do BTA (pH = 10,0) poderia, para a mesma reflexão, pode ser considerado o pico a $2\theta = 11,68^\circ$. Estas reflexões traduzir-se-iam em espaçamentos basais de ambas as tentativas de intercalação com BTA de $d(003) = 9,68$

Å, para a intercalação a pH = 8,1, e $d(003) = 9,73$ Å, para pH = 10,0. Assim, apesar de existir um pequeno aumento do espaçamento basal, a grande diferença de tamanho entre as moléculas, indica que, provavelmente, a intercalação não foi bem sucedida. Serdechnova *et al.* [43] obtiveram um espectro de DRX muito semelhante aos da Figura 18 e da Figura 19, tendo usado, mesmo assim, um pH diferente dos usados neste trabalho (pH = 11.5). Esses autores reproduziram a mesma reação, mantendo as mesmas condições, mas em vez de usarem HDL-NO₃ como material precursor, usaram óxido de zinco, e obtiveram um espectro de DRX semelhante ao resultante da troca iónica de HDL-NO₃ com BTA, e nas mesmas posições aos representados na Figura 18 e na Figura 19. Isto deve-se ao facto do BTA poder formar complexos metálicos com o zinco [44,45], e que podem dar origem às fases de (ZnO + BTA) apresentadas nas figuras de DRX. Os HDL são conhecidos por apresentarem alguma dissolução das camadas catiónicas em solução aquosa [1]. Devido a isto, a presença de BTA é capaz de reagir com Zn²⁺ em solução e, assim, formar complexos, podendo deslocar ainda mais o equilíbrio na direção da dissolução, e, assim, acabando por dissolver/decompor o HDL-NO₃ existente inicialmente em solução.

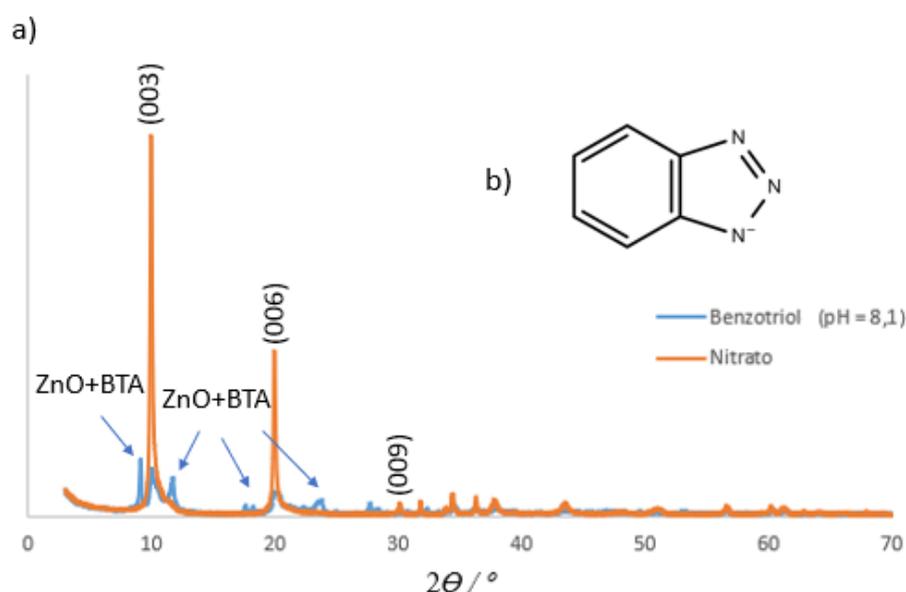


Figura 18 - a) DRX do HDL-NO₃ e do HDL-BTA (pH = 8,1); **b)** Molécula de BTA desprotonada.

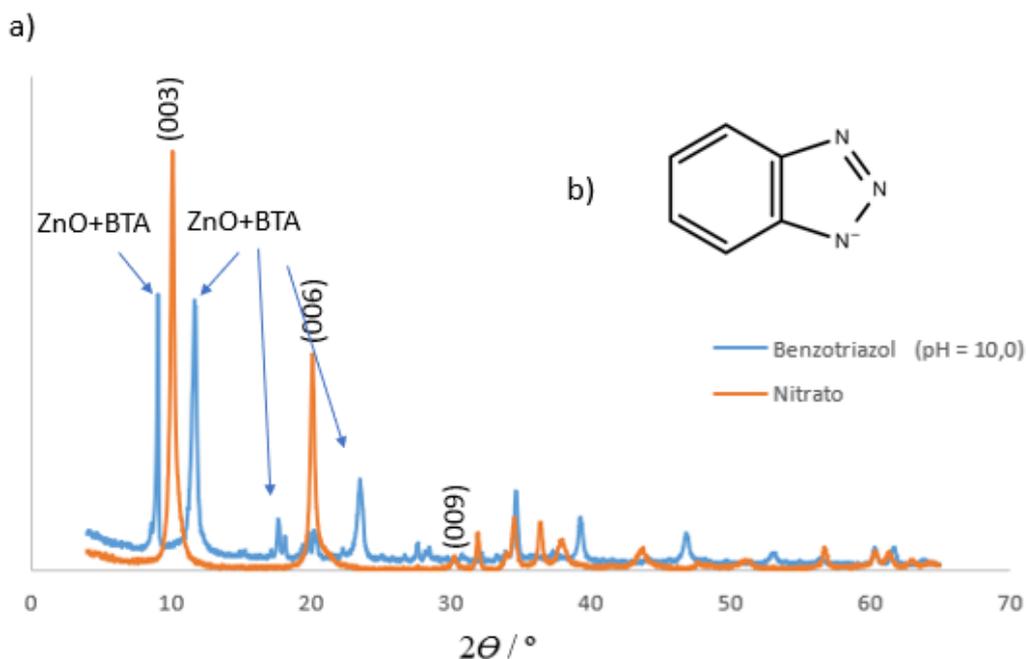


Figura 19 - a) DRX do HDL-NO₃ e do HDL-BTA (pH = 10,0); b) Molécula de BTA desprotonada.

4.2 - Análise dos resultados de intercalação parcial

Em alguns casos foi possível identificar uma intercalação parcial em que é possível verificar picos que podem ser associados ao composto intercalado, mas o nitrato continua presente em grandes quantidades, pelo que não foi considerada uma troca aniônica completa. Em análise temos os casos do ácido ascórbico com valores de pH = 8,0, 8,7 e 10,0, o ácido salicílico com um pH = 8,2 e o ácido sulfâmico com um pH = 8,0.

Em relação ao ácido ascórbico, apesar de se terem testado 3 pHs diferentes obtiveram-se apenas intercalações parciais, ou, pelo menos, que deixaram dúvidas. Os difratogramas relativos ao ácido ascórbico a pH = 8,0, 8,7 e 10,0 são apresentados nas Figura 20, Figura 21 e Figura 22, estando os picos de difração correspondentes às reflexões (003), (006) e (009) localizados a $2\theta = 9,89^\circ$, $19,9^\circ$ e $33,84^\circ$ (pH = 8,0), a $2\theta = 10,32^\circ$, $20,51^\circ$ e $34,59^\circ$ (pH = 8,7), e a 2θ foram observados a $10,7^\circ$, $21,62^\circ$ e $31,83^\circ$ (pH = 10,0), respetivamente. Com a posição dos picos do ácido ascórbico foi possível calcular os espaçamentos basais a pH = 8,0, 8,7 e 10,0 iguais a $8,93\text{Å}$, $8,56\text{Å}$ e $8,2\text{Å}$, respetivamente. É possível verificar que o espaçamento basal está muito próximo do nitrato ($d(003) = 8,85\text{Å}$), apesar do ácido ascórbico ser uma molécula com uma dimensão superior. A hipótese para se considerar uma intercalação parcial, é que o pH da solução

permite desprotonar o ácido ascórbico em vários pontos (Figura 23), permitindo contrabalançar a mesma molécula uma maior área de camada catiónica. Assim, existe mais espaço nas intercamadas para que o ácido ascórbico adote uma posição planar ou próxima de planar, quando co-intercalado com o anião nitrato e hidróxido [46], levando a um espaço, em camadas catiónicas, mais reduzido do que o que se esperava através da observação apenas do tamanho da molécula. Esta hipótese pode ter algum suporte através da relação observada entre pH e distância entre camadas, conforme apresentada na Figura 24.

Aisawa *et al.* [46], tentaram a intercalação do ácido ascórbico tendo obtido distâncias entre camadas semelhantes às verificadas neste trabalho, apontando a co-intercalação com os aniões nitrato e hidróxido, e a sua posição planar na intercamada, suportado por análise elementar, espectroscopia de infravermelho e ressonância magnética nuclear de estado sólido. Estes autores obtiveram distâncias entre camadas entre 8,4 Å e 10,6 Å, pelo que ajuda a suportar a intercalação parcial obtida neste trabalho. Por outro lado, estes autores também apontaram outras possíveis reações do ácido ascórbico, como, por exemplo, a sua oxidação quando exposto à luz ou calor, sendo que a sua intercalação no HDL ajuda a evitar a este processo [46], ou até a sua adsorção na superfície do HDL, que, mesmo assim, permite este material atuar por troca iónica através da troca na superfície.

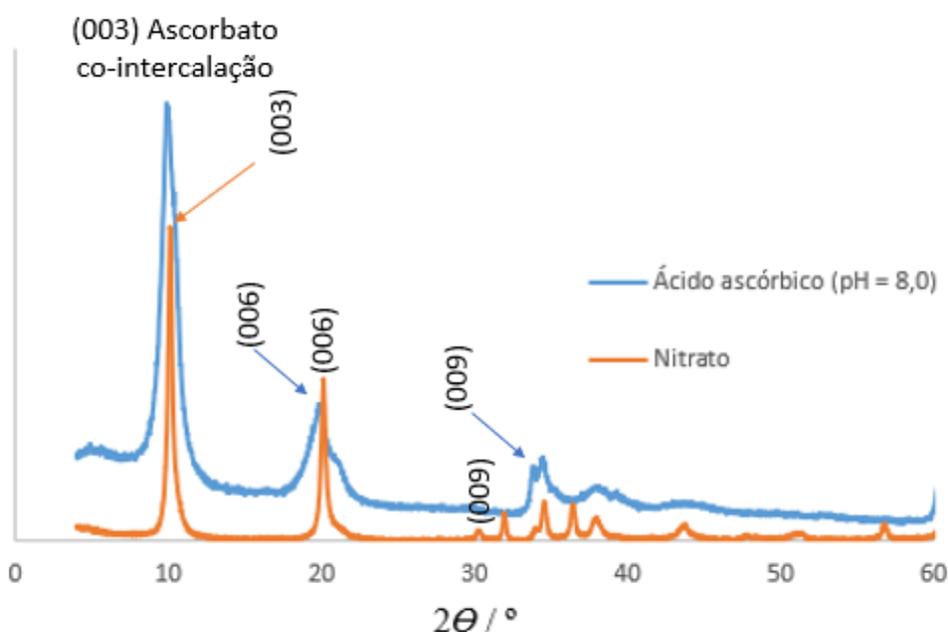


Figura 20 - a) DRX do HDL-NO₃ e HDL-ascorbato (pH = 8,0).

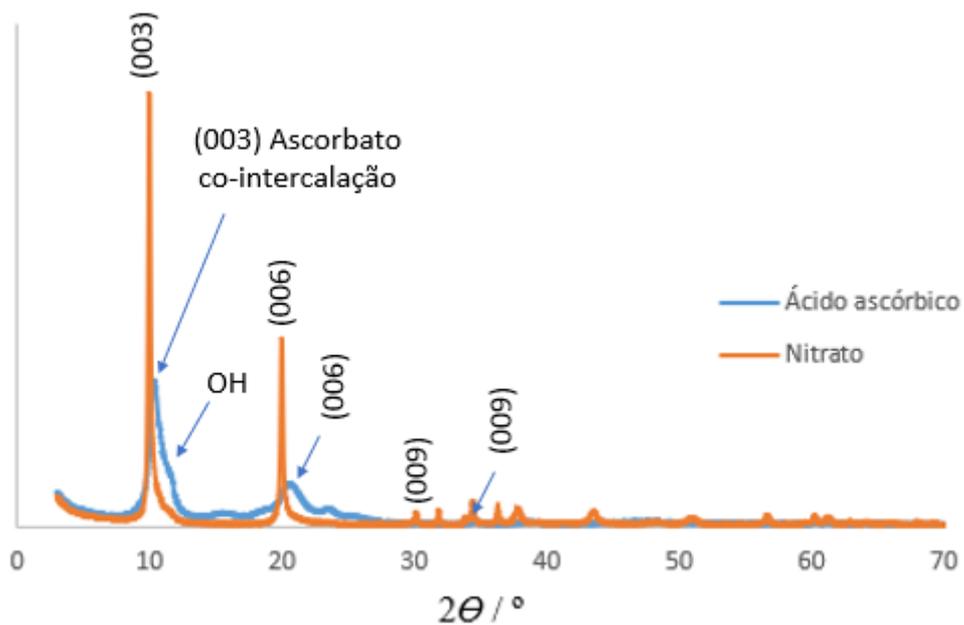


Figura 21 - a) Difratoograma de raios-X do HDL-NO₃ e HDL-ascorbato (pH = 8,7).

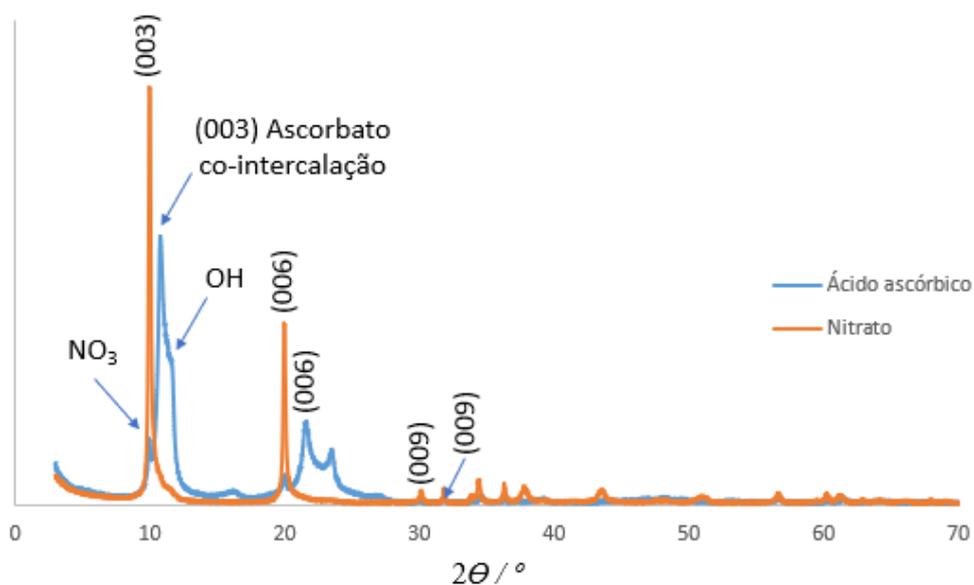


Figura 22 - a) Difratoograma de raios-X do HDL-NO₃ e HDL-ascorbato (pH = 10,0); b) Molécula ácido ascórbico desprotonada.

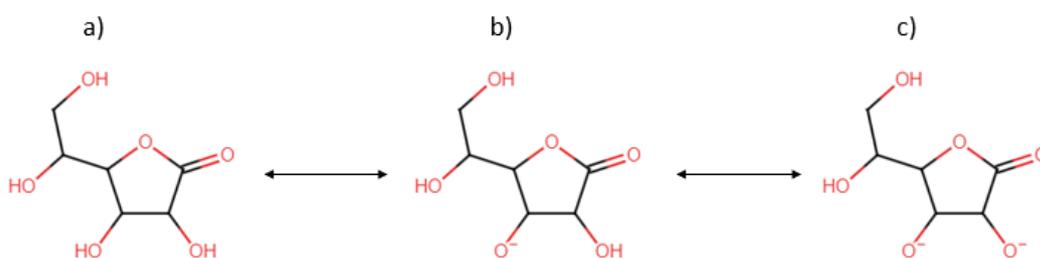


Figura 23 - a) Molécula de ácido ascórbico; b) Aniã da molécula de ácido ascórbico; c) Di-aniã da molécula de ácido ascórbico.

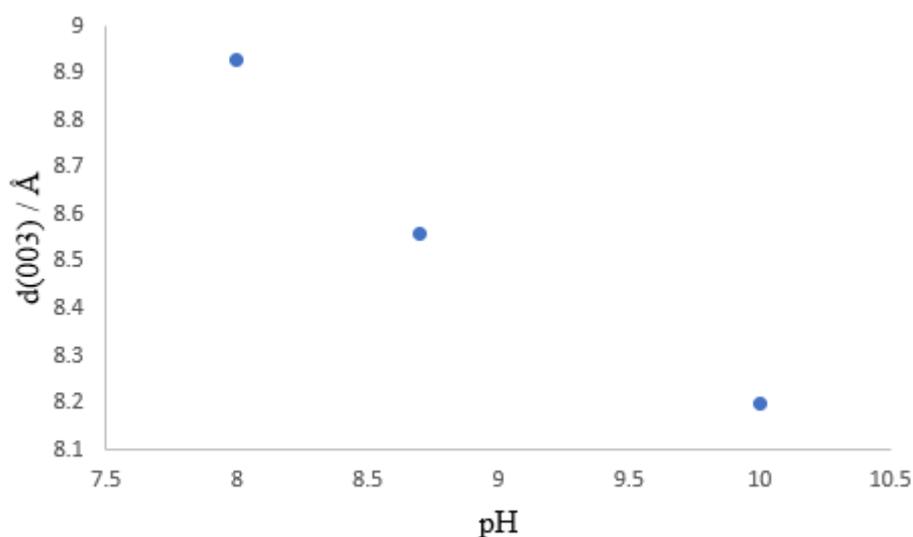


Figura 24 - Correlação entre pH e espaçamento basal do ácido ascórbico.

O difratograma do ácido salicílico (pH = 8,2), representado na Figura 25 apresenta picos localizados a $2\theta = 4,02^\circ$ e $8,08^\circ$, correspondentes a (003) e (006), respetivamente. No final calculou-se uma distância entre camadas catiónicas de ($d = 21,97 \text{ Å}$), ou seja, verificou-se um aumento do espaçamento entre camadas, sendo que pode haver uma co-intercalação com OH^- , o que se considerou como intercalação parcial.

O difratograma após a troca iónica com ácido salicílico a pH = 9,4 (Figura 26), levanta bastantes dúvidas sobre se existe um pico a ângulos reduzidos que possa indicar uma intercalação, por isso, foi considerado como não tendo sido intercalado.

Uma vez que para pHs inferiores a intercalação parece ser a mais bem-sucedida, como trabalho futuro, propõe-se a síntese a pH = 6.

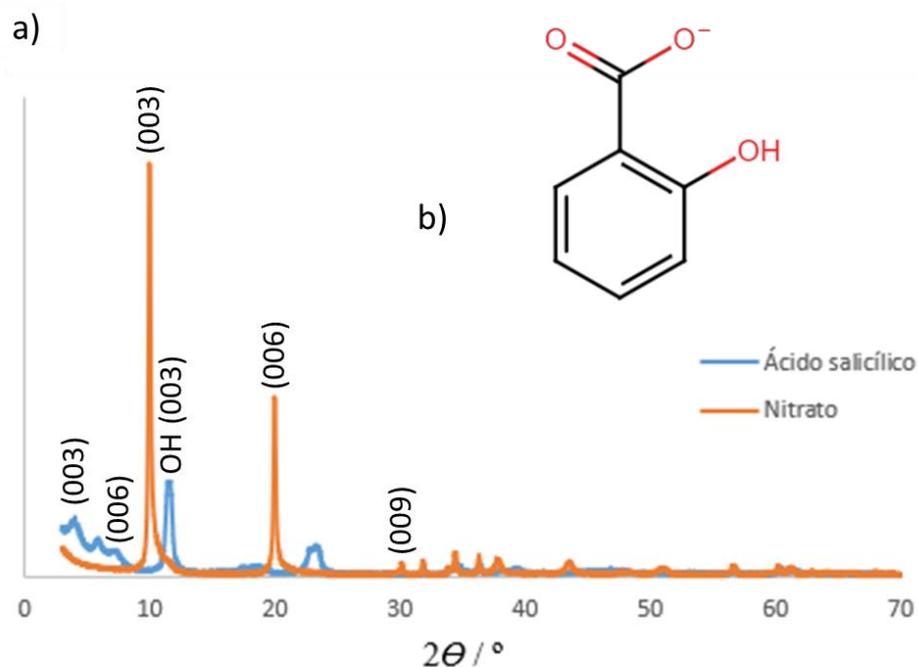


Figura 25 - a) Difratoograma de raios-X do HDL-NO₃ e HDL-ácido salicílico (pH = 8,2); b) Molécula ácido salicílico desprotonado.

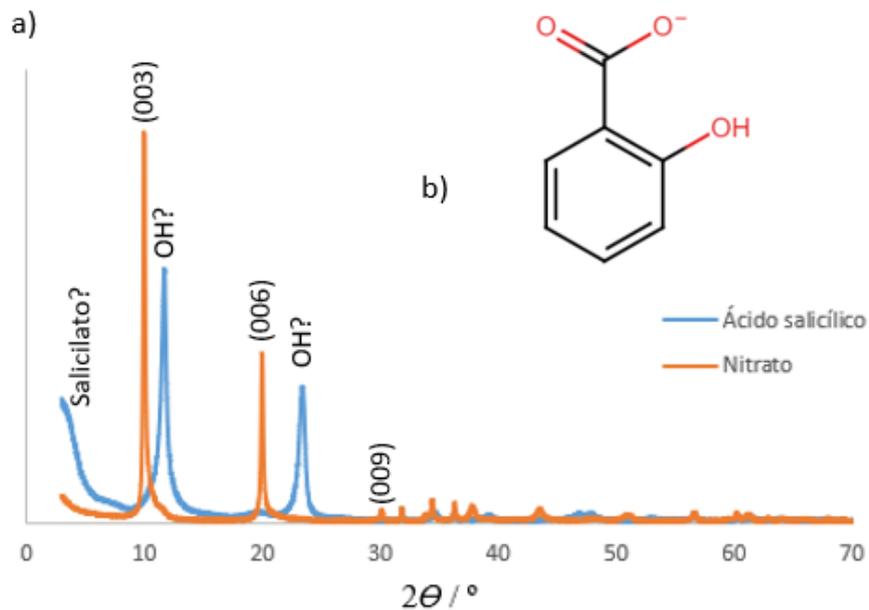


Figura 26 - a) Difratoograma de raios-X do HDL-NO₃ e HDL-ácido salicílico (pH = 9,4); b) Molécula ácido salicílico desprotonado.

O espectro do ácido sulfâmico com pH igual a 10 não intercalou, mas o espectro representado na Figura 27, a pH igual a 8,0, apresenta picos localizados a $2\theta = 10,65^\circ$,

21,24° e 31,91°, correspondendo respectivamente a (003), (006), (009). A Figura 28 representa a tentativa de intercalação a um pH= 10,2 onde não houve sucesso. Este difratograma apresenta, por um lado, uma co-intercalação como o nitrato, e, por outro lado, fases muito similares ao próprio difratograma do nitrato, com o espaçamento basal ligeiramente inferior ($d = 8,3 \text{ \AA}$). Isto acontece porque o ácido sulfâmico tem uma estrutura semelhante ao nitrato, mas com um NH_2 em vez de um O, o que vai causar menos repulsão electrostática, podendo até mesmo formar pontes de hidrogénio com os átomos de oxigénio das outras moléculas de ácido sulfâmico.

Assim, isto pode resultar numa diminuição ligeira do espaçamento entre camadas, o que poderá justificar ser um dos 3 picos que aparecem na mesma zona do pico $d(003)$, podendo corresponder o pico à esquerda ao nitrato, o pico do meio ao sulfamato e pico à direita ao OH^- . Como trabalho futuro e para se aproximar cada vez mais de uma intercalação completa, o pH da solução deverá ser ajustado a um valor próximo de 6.

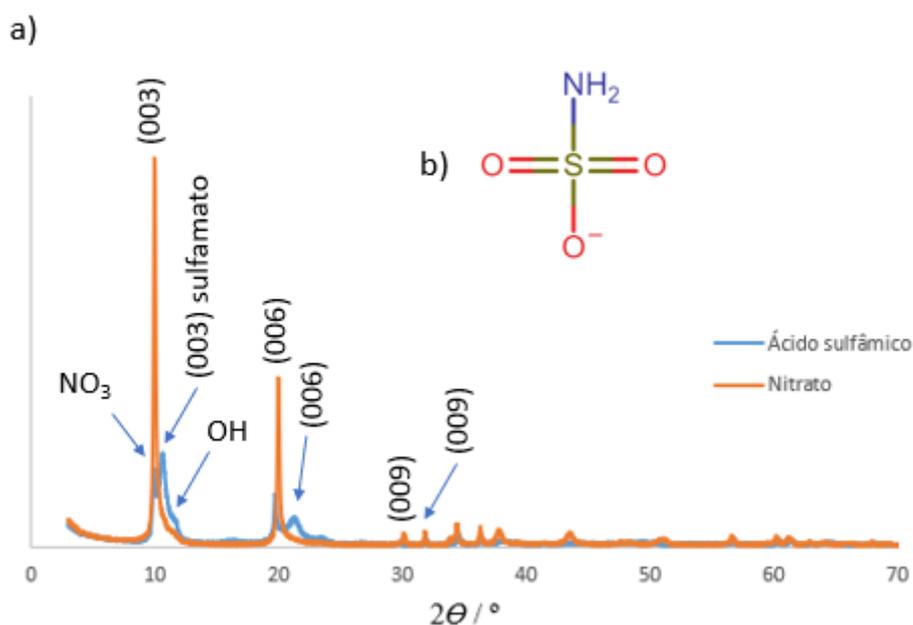


Figura 27 - a) Difratograma de raios-X do HDL- NO_3 e HDL-sulfamato (pH = 8,0); b) Molécula ácido sulfâmico desprotonado.

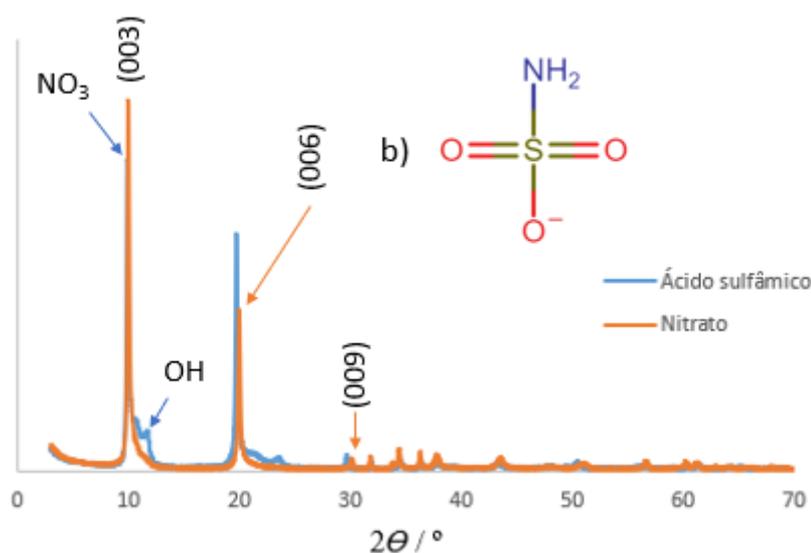


Figura 28 - a) Difratoograma de raios-X do HDL-NO₃ e HDL-sulfamato (pH = 10,2); b) Molécula ácido sulfâmico desprotonado.

Para a construção de modelos binários de classificação das intercalações, considerou-se as intercalações cujas moléculas intercalaram parcialmente, como sucesso, ou seja, para efeitos de uma avaliação virtual, é possível intercalar, embora a melhoria do DRX possa estar sujeita a mais testes como alguns dos pHs que foram propostos.

4.3 - Análise dos resultados de aprendizagem automática

Conforme indicado na secção 3, após o tratamento de dados, utilizou-se o método de eliminação recursiva de descritores para remover iterativamente descritores menos importantes, criando um subconjunto de descritores que maximiza a métrica estatística selecionada (CCM), tendo sido identificados os seguintes descritores como sendo os mais importantes: pH, n° de sínteses, pKa, pH – pKa, peso molecular, ponto de fusão, XLogP3, número de doadores/aceitadores de pontes de hidrogénio, número de ligações rotativa, área de superfície polar, número de átomos pesados e complexidade.

Para validar a seleção de descritores efetuada pelo método de eliminação recursiva de descritores, foi realizada a validação cruzada com todos os descritores disponíveis e também com todos aqueles que foram selecionados.

Na Tabela 5, são apresentados os resultados referentes à validação cruzada e ao teste do modelo. Através dos resultados obtidos, já é possível verificar que usando todos os descritores disponíveis se obtêm valores relativamente elevados para as várias métricas estatísticas. Contudo, com o modelo a excluir 4 descritores, ficando com apenas os 13

descritores selecionados, os parâmetros aumentam um pouco ou ficam iguais, embora dentro do desvio padrão obtido para a média das várias parcelas usadas na validação cruzada. Assim, válida a seleção efetuada anteriormente, ou seja, os descritores eliminados não são relevantes para o modelo.

Quanto aos valores em si, obtidos com validação cruzada com os descritores selecionados, pode verificar-se que a exatidão, exatidão balanceada, sensibilidade e precisão, são todos acima de 70 %, o que já permite prever corretamente a maioria das sínteses, e são todos semelhantes, o que indica que os casos de sucesso e insucesso são avaliados igualmente bem pelo modelo. Já os valores de Kappa e de CCM apontam para um acordo substancial do modelo, de acordo com a valores da Tabela 3 do subcapítulo 3.1.5.

Para o teste final é possível verificar na tabela que os resultados apontam para uma melhoria de todos os parâmetros, exceto na sensibilidade, provando assim que o modelo teve um bom desempenho para novos dados, que nunca foram usados na escolha dos descritores e otimização dos seus parâmetros. Os valores de Kappa e CCM são fatores importantes para um estudo de classificação, pois ajudam a encontrar erros na avaliação e interpretação dos resultados que podem não ser considerados pelas restantes métricas, por essa razão, como os valores obtidos no teste estiveram entre os 0,41 e 0,60, e verificando os valores da Tabela 3 do subcapítulo 3.1.5, conclui-se que os testes ao modelo tiveram um acordo substancial, mostrando que o modelo é viável para uma análise virtual de possíveis compostos a serem intercalados.

Tabela 5 - Resultados obtidos na validação e teste do modelo.

Validação/Teste	Exatidão / %	Exatidão Balanceada / %	Sensibilidade / %	Precisão / %	Kappa	CCM
Validação cruzada - Todos os descritores	73 ± 17	72 ± 15	76 ± 25	76 ± 13	0,45 ± 0,30	0,46 ± 0,32
Validação cruzada - Descritores selecionados	73 ± 17	73 ± 15	76 ± 25	78 ± 15	0,48 ± 0,30	0,50 ± 0,32
Teste final	75	83	66	100	0,50	0,58

4.4 - Importância dos descritores

O diagrama de caixa apresentado na Figura 29 mostra o resumo de cinco números de um conjunto de dados, sendo estes o mínimo, primeiro quartil, mediana, terceiro quartil, e máximo. Este gráfico mostra o resultado da importância dos descritores obtida através do método da permuta de descritores.

Quando a coluna correspondente a um determinado descritor é baralhada aleatoriamente, é usado um modelo desenvolvido anteriormente (florestas aleatórias com os descritores selecionados e parâmetros otimizados) para se prever os resultados. Se o valor da métrica estatística usada como figura de mérito deteriorar bastante, quer dizer que o descritor é importante. Se a métrica estatística permanecer igual, significa que o descritor é pouco importante.

Como é possível observar, o pH é o descritor com maior importância, sendo, na prática, a principal condição experimental que dita se determinado composto intercala ou não. Este aspeto é interessante, uma vez que mostra que o modelo consegue captar estatisticamente algo que é um aspeto intrínseco à química-física do processo de troca iónica.

Outros descritores considerados importantes pelo modelo foram características individuais das moléculas como o peso molecular, ponto de fusão, número de ligações rotativas, XLogP3, complexidade e área de superfície polar.

O descritor com menos importância é o número de passos de intercalação, uma vez que, com apenas um passo de intercalação, é possível verificar já a possibilidade de intercalação de uma determinada molécula. Um segundo passo, pode eventualmente ajudar a intercalação a ser mais completa. Todavia, foi usado apenas um passo de intercalação, tanto neste estágio curricular, como em intercalações prévias feitas na SMT de maneira a reduzir os recursos e acelerar o processo de desenvolvimento. No laboratório do SECOP da Universidade de Aveiro, utilizaram-se dois passos de intercalação, de modo a garantir a maior qualidade possível dos materiais para fins de publicação científica.

O modelo também atribui a mesma baixa diminuição do CCM a outros descritores, tais como pKa, pH – pKa, número de dadores/aceitadores de pontes de hidrogénio, e número de átomos pesados.

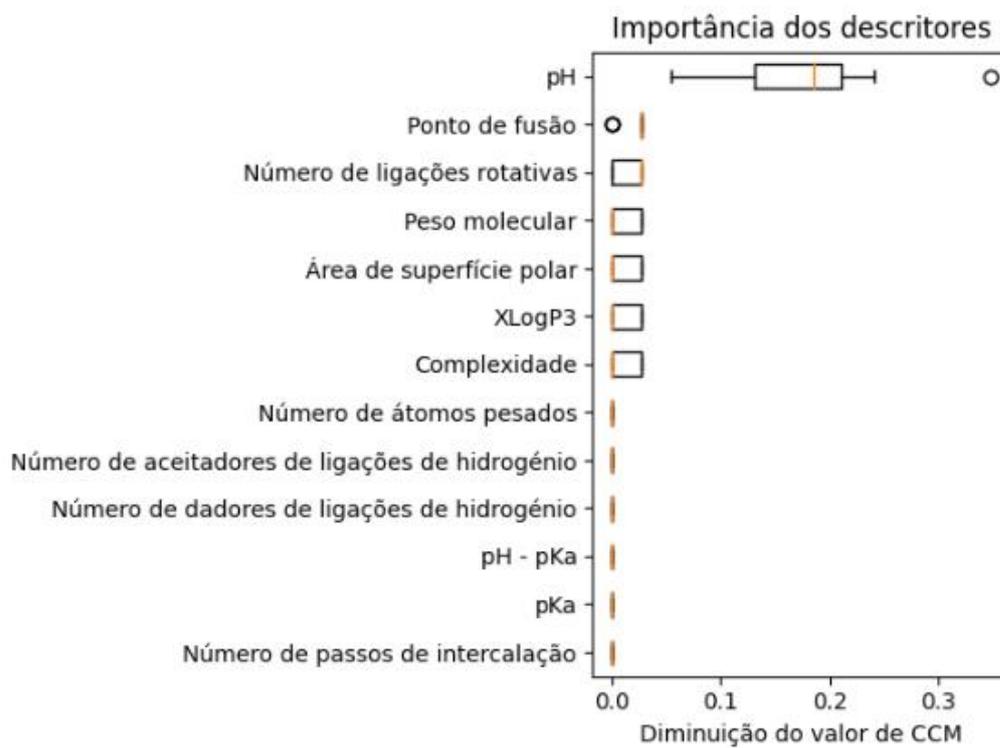


Figura 29 - Diagrama de importância de descritores avaliada pelo método da permuta de descritores.

Capítulo 5 - Conclusões e trabalho futuro

Neste trabalho híbrido experimental/computacional, recorreu-se a diferentes métodos de AA, para produzir um modelo que permita prever se um determinado composto funcional pode ser intercalado em materiais nanoestruturados baseados em HDL.

Estes materiais são produzidos a nível industrial pela empresa SMT e permitem o armazenamento e libertação controlada dos compostos intercalados consoante variação das condições do meio externo, permitindo, por exemplo, dar diferentes funcionalidades a estes materiais quando são incluídos como aditivos nanoestruturados em revestimentos.

Assim, com a aplicação de técnicas de AA aos resultados obtidos permitirá à empresa minimizar o tempo e recursos financeiros necessários para o desenvolvimento de novos produtos, como inibidores de corrosão, biocidas/antimicrobianos, sensores, antocianinas, fármacos e/ou tensoativos.

Os HDL pertencem a uma classe de materiais nanoestruturados sintéticos capazes de realizar troca iónica, que podem ser sintetizados de maneira relativamente barata e potencialmente reciclável.

O objetivo principal foi prever se um determinado composto consegue ou não ser intercalado em HDL. Para tal, foi preciso obter e organizar o maior número de dados possível, considerando dados obtidos experimentalmente, tanto neste trabalho, como outros obtidos anteriormente na UA e na SMT. Desta forma, uma grande quantidade de dados, tais como, 53 compostos, sendo para alguns foram testadas diferentes condições de síntese totalizando 97 sínteses, e, ainda, 17 descritores, contribuíram para obter um modelo mais robusto.

Com este conjunto de dados foi possível obter um modelo que considera diferentes tipos de condições experimentais, através de uma linguagem de programação de código aberto (Python), o que poderá possibilitar a construção de aplicações em trabalhos futuros.

O método das florestas aleatórias foi usado na construção do modelo, pois este algoritmo é uma combinação de um conjunto de árvores de decisão com o método agregação, obtendo uma previsão mais exata e estável.

Após o tratamento de dados, utilizou-se o método de eliminação recursiva de descritores para remover, iterativamente, descritores menos importantes, criando um subconjunto de descritores que maximiza a métrica estatística selecionada (CCM).

Para validar a seleção de descritores efetuada pelo método de eliminação recursiva de descritores, foi realizada a validação cruzada com todos os descritores disponíveis e também com todos aqueles que foram selecionados, sendo que foi validada a escolha de descritores pelo método de eliminação recursiva.

Através dos valores obtidos da validação cruzada com os descritores selecionados, verifica-se que a exatidão, exatidão balanceada, sensibilidade e precisão, estão todos acima de 70 %, o que já permite prever corretamente a maioria das sínteses, e os valores são semelhantes para todas as métricas, o que indica que os casos de sucesso e insucesso são avaliados igualmente bem pelo modelo.

Para o teste final foi possível verificar que os resultados apontam para uma melhoria de todos os parâmetros, exceto na sensibilidade, provando assim que o modelo teve um bom desempenho para novos dados, e que nunca foram usados na escolha dos descritores e otimização dos seus parâmetros.

Os valores de Kappa e CCM valores obtidos no teste estiveram entre os 0,41 e 0,60, podendo concluir-se que o teste ao modelo teve um acordo substancial, tornando o modelo viável.

O pH foi o descritor com maior importância, pois é este que dita, na maior parte dos casos, se o composto intercala ou não. A grande quantidade de intercalações onde a desprotonação é mais um fator para ajudar na intercalação ou não, enaltece mais o pH ser o descritor mais importante. O descritor com menor importância é o número de passos de intercalação, o que é algo que tinha sido verificado anteriormente na SMT para tornar as sínteses mais rápidas e acessíveis financeiramente.

No futuro, como proposta para dar continuação ao trabalho, de modo a criar um método de aprendizagem automática, capaz de prever se um composto é capaz de intercalar ou não, com maior confiança e capaz de desenhar efetivamente novos HDL funcionais, pode considerar-se aumentar a base de dados através do teste de diferentes moléculas, ou testar HDL com diferentes combinações de metais. Por outro lado, podem testar-se diferentes condições para os casos em que não foi verificada intercalação, ou esta foi apenas parcial, tais como diferentes pHs, diferentes concentrações iniciais dos compostos, aumentar o número de passos de intercalação, ou até aumentar a temperatura para tentar promover a troca-iônica. Ao nível do modelo de AA também é importante testar outros algoritmos (por exemplo, redes neuronais artificiais), de modo a encontrar um método mais exato para este tipo de trabalho. No final, espera-se gerar uma aplicação

para ser usada a nível empresarial, nomeadamente pela SMT, de forma a reduzir custos e ganhar tempo para desenvolvimento de novos produtos e de novos projetos.

Referências Bibliográficas

- [1] T.L.P. Galvão, C.S. Neves, A.P.F. Caetano, F. Maia, D. Mata, E. Malheiro, M.J. Ferreira, A.C. Bastos, A.N. Salak, J.R.B. Gomes, J. Tedim, M.G.S. Ferreira, Control of crystallite and particle size in the synthesis of layered double hydroxides: Macromolecular insights and a complementary modeling tool, *J Colloid Interface Sci.* 468 (2016) 86–94. <https://doi.org/10.1016/j.jcis.2016.01.038>.
- [2] X. Duan, D.G. Evans, *Layered Double Hydroxides*, Springer-Verlag, Berlin/Heidelberg, 2006. <https://doi.org/10.1007/b100426>.
- [3] J. Tedim, T.L.P. Galvão, K.A. Yasakau, A. Bastos, J.R.B. Gomes, M.G.S. Ferreira, Layered double hydroxides for corrosion-related applications—Main developments from 20 years of research at CICECO, *Front Chem.* 10 (2022). <https://doi.org/10.3389/fchem.2022.1048313>.
- [4] Y. Zhang, H. Xu, S. Lu, Preparation and application of layered double hydroxide nanosheets, *RSC Adv.* 11 (2021) 24254–24281. <https://doi.org/10.1039/d1ra03289e>.
- [5] F. Li, X. Duan, Applications of layered double hydroxides, *Struct Bond.* 119 (2005) 193–223. https://doi.org/10.1007/430_007.
- [6] H. Ye, S. Liu, D. Yu, X. Zhou, L. Qin, C. Lai, F. Qin, M. Zhang, W. Chen, W. Chen, L. Xiang, Regeneration mechanism, modification strategy, and environment application of layered double hydroxides: Insights based on memory effect, *Coord Chem Rev.* 450 (2022) 214253. <https://doi.org/10.1016/j.ccr.2021.214253>.
- [7] L. Jin, X. Zhou, F. Wang, X. Ning, Y. Wen, B. Song, C. Yang, D. Wu, X. Ke, L. Peng, Insights into memory effect mechanisms of layered double hydroxides with solid-state NMR spectroscopy, *Nat Commun.* 13 (2022) 6093. <https://doi.org/10.1038/s41467-022-33912-7>.
- [8] D. Tonelli, E. Scavetta, M. Giorgetti, Layered-double-hydroxide-modified electrodes: electroanalytical applications, *Anal Bioanal Chem.* 405 (2013) 603–614. <https://doi.org/10.1007/s00216-012-6586-2>.
- [9] A. Vaccari, Preparation and catalytic properties of cationic and anionic clays, *Catal Today.* 41 (1998) 53–71. [https://doi.org/10.1016/S0920-5861\(98\)00038-8](https://doi.org/10.1016/S0920-5861(98)00038-8).
- [10] A. Vaccari, Clays and catalysis: a promising future, *Appl Clay Sci.* 14 (1999) 161–198. [https://doi.org/10.1016/S0169-1317\(98\)00058-1](https://doi.org/10.1016/S0169-1317(98)00058-1).

- [11] A.C. Bouali, M. Serdechnova, C. Blawert, J. Tedim, M.G.S. Ferreira, M.L. Zheludkevich, Layered double hydroxides (LDHs) as functional materials for the corrosion protection of aluminum alloys: A review, *Appl Mater Today*. 21 (2020) 100857. <https://doi.org/10.1016/j.apmt.2020.100857>.
- [12] M.L. Zheludkevich, J. Tedim, M.G.S. Ferreira, “Smart” coatings for active corrosion protection based on multi-functional micro and nanocontainers, in: *Electrochim Acta*, 2012: pp. 314–323. <https://doi.org/10.1016/j.electacta.2012.04.095>.
- [13] J. Tedim, A.C. Bastos, S. Kallip, M.L. Zheludkevich, M.G.S. Ferreira, Corrosion protection of AA2024-T3 by LDH conversion films. Analysis of SVET results, *Electrochim Acta*. 210 (2016) 215–224. <https://doi.org/10.1016/J.ELECTACTA.2016.05.134>.
- [14] M.L. Zheludkevich, S.K. Poznyak, L.M. Rodrigues, D. Raps, T. Hack, L.F. Dick, T. Nunes, M.G.S. Ferreira, Active protection coatings with layered double hydroxide nanocontainers of corrosion inhibitor, *Corros Sci*. 52 (2010) 602–611. <https://doi.org/10.1016/J.CORSCI.2009.10.020>.
- [15] R.G. Buchheit, H. Guan, S. Mahajanam, F. Wong, Active corrosion protection and corrosion sensing in chromate-free organic coatings, *Prog Org Coat*. 47 (2003) 174–182. <https://doi.org/10.1016/J.PORGOAT.2003.08.003>.
- [16] W. Zhang, B. Hurley, R.G. Buchheit, Characterization of Chromate Conversion Coating Formation and Breakdown Using Electrode Arrays, *J Electrochem Soc*. 149 (2002) B357. <https://doi.org/10.1149/1.1485774>.
- [17] H.N. McMurray, G. Williams, Inhibition of Filiform Corrosion on Organic-Coated Aluminum Alloy by Hydrotalcite-Like Anion-Exchange Pigments, *Corrosion*. 60 (2004) 219–228. <https://doi.org/10.5006/1.3287724>.
- [18] G. Williams, H.N. McMurray, Inhibition of Filiform Corrosion on Polymer Coated AA2024-T3 by Hydrotalcite-Like Pigments Incorporating Organic Anions, *Electrochemical and Solid-State Letters*. 7 (2004) B13. <https://doi.org/10.1149/1.1691529>.
- [19] R.L. Twite, G.P. Bierwagen, Review of alternatives to chromate for corrosion protection of aluminum aerospace alloys, *Prog Org Coat*. 33 (1998) 91–100. [https://doi.org/10.1016/S0300-9440\(98\)00015-0](https://doi.org/10.1016/S0300-9440(98)00015-0).
- [20] Y. Liu, P. Visser, X. Zhou, S.B. Lyon, T. Hashimoto, M. Curioni, A. Gholinia, G.E. Thompson, G. Smyth, S.R. Gibbon, D. Graham, J.M.C. Mol, H. Terryn,

- Protective Film Formation on AA2024-T3 Aluminum Alloy by Leaching of Lithium Carbonate from an Organic Coating, *J Electrochem Soc.* 163 (2016) C45–C53. <https://doi.org/10.1149/2.0021603jes>.
- [21] S.K. Poznyak, J. Tedim, L.M. Rodrigues, A.N. Salak, M.L. Zheludkevich, L.F.P. Dick, M.G.S. Ferreira, Novel Inorganic Host Layered Double Hydroxides Intercalated with Guest Organic Inhibitors for Anticorrosion Applications, *ACS Appl Mater Interfaces.* 1 (2009) 2353–2362. <https://doi.org/10.1021/am900495r>.
- [22] J. Tedim, A. Kuznetsova, A.N. Salak, F. Montemor, D. Snihirova, M. Pilz, M.L. Zheludkevich, M.G.S. Ferreira, Zn–Al layered double hydroxides as chloride nanotraps in active protective coatings, *Corros Sci.* 55 (2012) 1–4. <https://doi.org/10.1016/j.corsci.2011.10.003>.
- [23] T.L.P. Galvão, M. Wilhelm, J.R.B. Gomes, J. Tedim, Emerging trends in smart nanocontainers for corrosion applications, *Smart Nanocontainers: Micro and Nano Technologies.* (2020) 385–398. <https://doi.org/10.1016/B978-0-12-816770-0.00022-8>.
- [24] J. Tedim, S.K. Poznyak, A. Kuznetsova, D. Raps, T. Hack, M.L. Zheludkevich, M.G.S. Ferreira, Enhancement of Active Corrosion Protection via Combination of Inhibitor-Loaded Nanocontainers, *ACS Appl Mater Interfaces.* 2 (2010) 1528–1535. <https://doi.org/10.1021/am100174t>.
- [25] J. Tedim, Immobilization of active molecules in nanostructured materials for corrosion sensing, *NACE Corrosion.* (2016).
- [26] G. Pérez-Sánchez, T.L.P. Galvão, J. Tedim, J.R.B. Gomes, A molecular dynamics framework to explore the structure and dynamics of layered double hydroxides, *Appl Clay Sci.* 163 (2018) 164–177. <https://doi.org/10.1016/j.clay.2018.06.037>.
- [27] A.N. Salak, J. Tedim, A.I. Kuznetsova, J.L. Ribeiro, L.G. Vieira, M.L. Zheludkevich, M.G.S. Ferreira, Comparative X-ray diffraction and infrared spectroscopy study of Zn–Al layered double hydroxides: Vanadate vs nitrate, *Chem Phys.* 397 (2012) 102–108. <https://doi.org/10.1016/j.chemphys.2012.01.026>.
- [28] A. V. Radha, P.V. Kamath, C. Shivakumara, Conservation of order, disorder, and “crystallinity” during anion-exchange reactions among Layered Double Hydroxides (LDHs) of Zn with Al, *Journal of Physical Chemistry B.* 111 (2007) 3411–3418.

- https://doi.org/10.1021/JP0684170/SUPPL_FILE/JP0684170SI20061207_050603.PDF.
- [29] Kaitlin Tyler, K. Tyler, Understanding Machine Learning for Materials Science Technology, (2019). <https://www.ansys.com/blog/machine-learning-materials-science> (accessed June 19, 2023).
- [30] G. Shmueli, To Explain or to Predict?, *Statistical Science*. 25 (2010). <https://doi.org/10.1214/10-STS330>.
- [31] P.Wong, Predicting vs. Explaining, *Towards Data Science.*, (2019). <https://towardsdatascience.com/predicting-vs-explaining-69b516f90796> (accessed June 19, 2023).
- [32] B. Hu, K. Lu, Q. Zhang, X. Ji, W. Lu, Data mining assisted materials design of layered double hydroxide with desired specific surface area, *Comput Mater Sci*. 136 (2017) 29–35. <https://doi.org/10.1016/j.commatsci.2017.03.027>.
- [33] M.E., & S.A. Tipping, Sparse Bayesian learning and the relevance vector machine, *Journal of Machine Learning Research*. 1 (2001) 211–244. <https://doi.org/10.1162/15324430152748236>.
- [34] Q. Zhang, X. Zhai, P. Xiong, L. Kou, X. Ji, W. Lu, Prediction and synthesis of novel layered double hydroxide with desired basal spacing based on relevance vector machine, *Mater Res Bull*. 93 (2017) 123–129. <https://doi.org/10.1016/j.materresbull.2017.03.045>.
- [35] K. Lu, Machine Learning Aided Discovery of the Layered Double Hydroxides with the Largest Basal Spacing for Super-Capacitors, *Int J Electrochem Sci*. (2021). <https://doi.org/10.20964/2021.11.22>.
- [36] ChemBk, (2015). <https://www.chembk.com/en> (accessed August 25, 2023).
- [37] Drugbank, (2006). <https://go.drugbank.com/> (accessed August 25, 2023).
- [38] Pubchem, (2004). <https://pubchem.ncbi.nlm.nih.gov/> (accessed August 25, 2023).
- [39] W. McKinney, Data Structures for Statistical Computing in Python, in: 2010: pp. 56–61. <https://doi.org/10.25080/Majora-92bf1922-00a>.
- [40] Tiago L. P. Galvão, A machine learning approach for the prediction of optical conversion efficiency in luminescent solar concentrators, (Em preparação), (2023).
- [41] T.L.P. Galvão, G. Novell-Leruth, A. Kuznetsova, J. Tedim, J.R.B. Gomes, Elucidating Structure–Property Relationships in Aluminum Alloy Corrosion Inhibitors by Machine Learning, *Journal of Physical Chemistry C*. 124 (2020) 5624–5635. <https://doi.org/10.1021/acs.jpcc.9b09538>.

- [42] Inês Manuela Alves Ferreira, Seleção de inibidores de corrosão orgânicos para aplicações aeronáuticas através de Aprendizagem Automática, Universidade de Aveiro, 2021.
- [43] M. Serdechnova, A.N. Salak, F.S. Barbosa, D.E.L. Vieira, J. Tedim, M.L. Zheludkevich, M.G.S. Ferreira, Interlayer intercalation and arrangement of 2-mercaptobenzothiazolate and 1,2,3-benzotriazolate anions in layered double hydroxides: In situ X-ray diffraction study, *J Solid State Chem.* 233 (2016) 158–165. <https://doi.org/10.1016/j.jssc.2015.10.023>.
- [44] L. Li, X. Wang, G. Shen, R. Wang, D. Shen, Synthesis and crystal structure of $Zn[(C_6H_5N_3)_4(H_2O)_2](ClO_4)_2$ and $Cu[(C_6H_5N_3)_4(H_2O)_2](ClO_4)_2$, *Inorg Chem Commun.* 5 (2002) 453–457. [https://doi.org/10.1016/S1387-7003\(02\)00441-0](https://doi.org/10.1016/S1387-7003(02)00441-0).
- [45] T. Kosec, D.K. Merl, I. Milošev, Impedance and XPS study of benzotriazole films formed on copper, copper–zinc alloys and zinc in chloride solution, *Corros Sci.* 50 (2008) 1987–1997. <https://doi.org/10.1016/j.corsci.2008.04.016>.
- [46] S. Aisawa, N. Higashiyama, S. Takahashi, H. Hirahara, D. Ikematsu, H. Kondo, H. Nakayama, E. Narita, Intercalation behavior of l-ascorbic acid into layered double hydroxides, *Appl Clay Sci.* 35 (2007) 146–154. <https://doi.org/10.1016/j.clay.2006.09.003>.